

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 27, 2004, 03:11:33 ; Search time 4007.03 Seconds
(without alignments)
10708.573 Million cell updates/sec
Title: US-10-658-691-3
Perfect score: 990
Sequence: 1 atgtaccactcaaatgta.....atttgacgcattttcttaa 990
Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.in.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.man.*
37: em.htg.vrt.*
38: em.sy.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*
Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	990	100.0	990	6	E17299
2	990	100.0	6881	1	D89064
3	990	100.0	44266	1	AB047548
4	990	100.0	199035	1	AB088420
5	988.4	99.8	6880	1	AB001723
6	871.6	88.0	20017	1	AB095952
7	130.8	13.2	8620	1	AF065891
8	117.2	11.8	5337	1	D83068
9	117.2	11.8	5337	6	E32813
10	109	11.0	4729	1	AF282898
11	109	11.0	4993	1	PF004052
12	107.8	10.9	5743	6	E39081
13	104.6	10.6	24448	1	AB024945
14	91.6	9.3	8154	1	AF157565
15	85	8.5	10580	1	AB102786
16	81.6	8.2	6658	1	AF001356
17	81.2	8.2	1026	1	FSH38289
18	81.2	8.2	10470	1	PST5663
19	73.8	7.5	4911	1	XSP012090
20	64.6	6.5	1180	1	PSETMOF
21	64.6	6.5	4729	6	AR362443
22	62.6	6.3	8456	1	RSP249207
23	56.6	5.7	7509	1	AF091716
24	54.6	5.5	11514	1	AF039533
25	54.6	5.5	12808	1	PSEORF1
26	53	5.4	6739	1	AY125981
27	53	5.4	9706	1	AF010471
28	53	5.4	83042	1	AF491307
29	53	5.4	101858	1	AY208917
30	50.2	5.1	1105	1	PAU86603
31	49.8	5.0	4355	1	FSENPADONA
32	45.6	4.6	1058	1	AY048766
33	45.6	4.6	14462	1	AB004059
34	43	4.3	2000	6	AX655393
35	42.4	4.3	346510	1	AP003011
36	41	4.1	4799	1	AB098505
37	40.8	4.1	6652	1	AF360864
38	40.4	4.1	2000	6	AX655393
39	39.6	4.0	10506	1	AE008048
40	39.6	4.0	11109	1	AE009082
41	39.2	4.0	9783	1	ACA564846
42	39	3.9	4721	6	E32624
43	39	3.9	7823	1	AB024741
44	38.8	3.9	125020	9	AF429315
45	38.6	3.9	6298	1	AB006479

ALIGNMENTS

RESULT 1
E17299
LOCUS
DEFINITION
E17299
gDNA encoding oxidase which converts dibenzo-p-dioxin to a diphenyl ether triol.
ACCESSION
E17299
VERSION
1 GI:5711982
KEYWORDS
JP 1998257895-A/8.
SOURCE
Pseudomonas sp.
ORGANISM
Bacteria; Proteobacteria.
REFERENCE
1 (bases 1 to 990)
AUTHORS
Omori, T. and Takami, K.
TITLE
OXIDASE GENE ORIGINATING FROM MICROORGANISM AND REMOVAL OF DIOXIN WITH THE SAME
E17299 990 bp DNA linear PAT 28-JUL-1999
gDNA encoding oxidase which converts dibenzo-p-dioxin to a diphenyl ether triol.

JOURNAL Patent: JP 1998257995-A 8 29-SEP-1998;
 ASahi CHEM IND CO., LTD.
 COMMENT Pseudomonas sp.
 PN JP 198257895-A/8
 PD 29-SEP-1998
 PF 18-MAR-1997 JP 1997084401
 PI OMORI TOSHIO, TAKAMI KAZUTAKA
 PC C12N15/09, A62D3/00, B09C1/10, C02F3/34, C07H21/04, PC
 C12N1/21/1/C12N9/02,
 PC (C12N15/09, C12R1:38), (C12N1/21, C12R1:19), (C12N9/02, C12R1:19);
 CC strandedness: Double;
 CC topology: linear;
 CC hypothetical: No;
 CC anti-sense: No; Location/Qualifiers
 FH Key
 FT source 1..990
 FT /organism="Pseudomonas sp."
 FT /strain="CA10"
 FT Location/Qualifiers
 FT 1..990
 FT /organism="Pseudomonas sp."
 FT /mol_type="genomic DNA"
 FT /db_xref="taxon:306"
 FEATURES
 source
 ORIGIN
 Query Match 100.0%; Score 990; DB 6; Length 990;
 Best Local Similarity 100.0%; Pred. No. 1.3e-256;
 Matches 990; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGTACCAACTCAAAATTTGAAGGCGAAGCGGACCTGCGGCTCAGGGAAGAGCGCTG 60
 Db 1 ATGTACCAACTCAAAATTTGAAGGCGAAGCGGACCTGCGGCTCAGGGAAGAGCGCTG 60
 QY 61 TTGGTCTCAGACACTGCTTAATGGTATCGATTTCCGTACGAGTGTGATCGGAGGTTC 120
 Db 61 TTGGTCTCAGACACTGCTTAATGGTATCGATTTCCGTACGAGTGTGATCGGAGGTTC 120
 QY 121 CGAGTATCAAAATTCGAGTCTACTCGAAGGGAATGCCAATCAATGTGGCGGATGCTCCA 180
 Db 121 CGAGTATCAAAATTCGAGTCTACTCGAAGGGAATGCCAATCAATGTGGCGGATGCTCCA 180
 QY 181 GGACTTTCTCGGAGATCGTGAGAGGGCAACGCCCATCTTGCATGCCAGTGGTGGCG 240
 Db 181 GGACTTTCTCGGAGATCGTGAGAGGGCAACGCCCATCTTGCATGCCAGTGGTGGCG 240
 QY 241 CTCTCAGACCTCGGATCAAAAGTCCAGTGCAGGACAGTAGTACGTCCTCAAGGATTCGAATC 300
 Db 241 CTCTCAGACCTCGGATCAAAAGTCCAGTGCAGGACAGTAGTACGTCCTCAAGGATTCGAATC 300
 QY 301 TCAAGATGGAAGCGGAAGTTGTTGAGTCCGGGCGCTTAATCAATGACCTGCTGTCCGTG 360
 Db 301 TCAAGATGGAAGCGGAAGTTGTTGAGTCCGGGCGCTTAATCAATGACCTGCTGTCCGTG 360
 QY 361 CGATTACGCACTGATGGCGGACGAAATTTCTCCCGGCGGAGTCTGCCTAGTAGAGGCA 420
 Db 361 CGATTACGCACTGATGGCGGCGGAAATTTCTCCCGGCGGAGTCTGCCTAGTAGAGGCA 420
 QY 421 GAGCAGTTCGACGCGCTGTTTCGCGCATATTCATGCGCAATTTAAAGAACCCCGAAGGC 480
 Db 421 GAGCAGTTCGACGCGCTGTTTCGCGCATATTCATGCGCAATTTAAAGAACCCCGAAGGC 480
 QY 481 ATATGGAGTCTTATTAAGAGGTTACCCAGAGGATTTAGTCTTGGCTTTTCGAA 540
 Db 481 ATATGGAGTCTTATTAAGAGGTTACCCAGAGGATTTAGTCTTGGCTTTTCGAA 540
 QY 541 AATAGAAAGAGGCGCTGCTCTATTTTTCGCGGACCAATGGGCAACATCTTTCTCCGT 600
 Db 541 AATAGAAAGAGGCGCTGCTCTATTTTTCGCGGACCAATGGGCAACATCTTTCTCCGT 600
 QY 601 CCAGGACCGGCGGAGGCTTTGATTTGGCGGCGTGCAGGCTCTGTATGCGGCC 660
 Db 601 CCAGGACCGGCGGAGGCTTTGATTTGGCGGCGTGCAGGCTCTGTATGCGGCC 660

QY 661 GCTATTTCACGCGCCTCGATCGCGAACAACAGACAGCCGTAAGATTTGTTCTACGGCTCA 720
 Db 661 GCTATTTCACGCGCCTCGATCGCGAACAACAGACAGCCGTAAGATTTGTTCTACGGCTCA 720
 QY 721 AGAATCTCCGCGGACGCTGTTTCGTTGATCGATATCGATCGATCGATCGATCGATCGATCG 780
 Db 721 AGAATCTCCGCGGACGCTGTTTCGTTGATCGATATCGATCGATCGATCGATCGATCGATCG 780
 QY 781 GTCTGCCAGGCGATTTACGGAACACGATAGCCTTTGCAAGGCGCCACCTGTTTATT 840
 Db 781 GTCTGCCAGGCGATTTACGGAACACGATAGCCTTTGCAAGGCGCCACCTGTTTATT 840
 QY 841 CATCAGGTTTCGACGCGGCTGTTGAAACCCCTACCGGAATACGAAATTTATCTTCC 900
 Db 841 CATCAGGTTTCGACGCGGCTGTTGAAACCCCTACCGGAATACGAAATTTATCTTCC 900
 QY 901 GCTCACCGCCTATGTCGACGCTACTCTCGTATGCTGCTCGGCAAGGTTTCCACGC 960
 Db 901 GCTCACCGCCTATGTCGACGCTACTCTCGTATGCTGCTCGGCAAGGTTTCCACGC 960
 QY 961 GATCAAAATTCATTTTACGCAATTTTCTAA 990
 Db 961 GATCAAAATTCATTTTACGCAATTTTCTAA 990
 RESULT 2
 D89064
 LOCUS 6881 bp DNA linear BCT 12-SEP-1997
 DEFINITION Pseudomonas sp. gene for carbazole 1,9a-dioxygenase, meta-cleavage
 enzyme, meta-cleavage compound hydrolase, complete cds.
 ACCESSION D89064
 VERSION D89064.1 GI:2317677
 KEYWORDS reductase component of carbazole 1,9a-dioxygenase; ferredoxin
 component of carbazole 1,9a-dioxygenase; meta-cleavage compound
 hydrolase; meta-cleavage enzyme; terminal dioxygenase component of
 carbazole 1,9a-dioxygenase.
 SOURCE Pseudomonas sp.
 ORGANISM Pseudomonas sp.
 Bacteria; Proteobacteria.
 REFERENCE 1 (sites)
 AUTHORS Sato, S.I., Ochiyama, N., Kimura, T., Nojiri, H., Yamane, H. and
 Omori, T.
 TITLE Cloning of genes involved in carbazole degradation of Pseudomonas
 sp. strain CA10: nucleotide sequences of genes and characterization
 of meta-cleavage enzymes and hydrolase
 J. Bacteriol. 179 (15), 4841-4849 (1997)
 JOURNAL 97386424
 MEDLINE 9244273
 PUBMED
 REFERENCE 2 (sites)
 AUTHORS Sato, S.I., Nam, J.W., Kasuga, K., Nojiri, H., Yamane, H. and Omori, T.
 TITLE Identification and characterization of genes encoding carbazole
 1,9a-dioxygenase in Pseudomonas sp. strain CA10
 J. Bacteriol. 179 (15), 4850-4858 (1997)
 JOURNAL 97386425
 MEDLINE 9244274
 PUBMED
 REFERENCE 3 (bases 1 to 6881)
 AUTHORS Omori, T.
 TITLE Direct Submission
 JOURNAL Submitted (14-NOV-1996) Toshio Omori, The University of Tokyo,
 Biotechnology Research Center; Yayoi 1-1-1, Bunkyo-ku, Tokyo 113,
 Japan [E-mail: assigyo@hongo.ecc.u-tokyo.ac.jp,
 Tel:03-3812-2111 (ex.3067), Fax:03-5802-3326]
 FEATURES
 source
 1..6881
 /organism="Pseudomonas sp."
 /mol_type="genomic DNA"
 /strain="CA10"
 /db_xref="taxon:306"
 572..1726
 /gene="carAa"
 572..1726
 /gene="carAa"

```

/codon_start=1
/transl_table=11
/product="terminal dioxygenase component of carbazole
1, 9a-dioxygenase"
/protein_id="BAA21728.1"
/db_xref="GI:2317678"
/translation="WANDDEAILKRVKGNAPYVDALQFRNHWTFWMSKEIDGEPK
TLKLGELLNVRIDGKLYCLDRCLHRGVQSVKVECKTSTTCWHAWTYRWDG
VLCDLTNPSTAQIGRQKLTYPQOAKGCVFIYLGDDGPPPLARDTPPNFLDDMEI
LGKQI IKSNRWLA VENGDFDP SHIYHKDSILVKNDLALPLGAPGDRKQOQTRVVD
DDVGRKGVYDLIGHGVPVPEGTIGGEVREGAYGEKIVANDISIMLPGLVKYNPPF
NPDMMQFWEVPI DENTHYFOTLGKPCANDEERKNYBOEPESKWKWALGFPNDDI
WAREAMVDFYADDKGNWNEILLFEVDEAIVARKLASEHNQGIQTOAHVSG"
1838..2992
/gene="carAb"
1838..2992
/gene="carAb"
/codon_start=1
/transl_table=11
/product="terminal dioxygenase component of carbazole
1, 9a-dioxygenase"
/protein_id="BAA21729.1"
/db_xref="GI:2317679"
/translation="WANDDEAILKRVKGNAPYVDALQFRNHWTFWMSKEIDGEPK
TLKLGELLNVRIDGKLYCLDRCLHRGVQSVKVECKTSTTCWHAWTYRWDG
VLCDLTNPSTAQIGRQKLTYPQOAKGCVFIYLGDDGPPPLARDTPPNFLDDMEI
LGKQI IKSNRWLA VENGDFDP SHIYHKDSILVKNDLALPLGAPGDRKQOQTRVVD
DDVGRKGVYDLIGHGVPVPEGTIGGEVREGAYGEKIVANDISIMLPGLVKYNPPF
NPDMMQFWEVPI DENTHYFOTLGKPCANDEERKNYBOEPESKWKWALGFPNDDI
WAREAMVDFYADDKGNWNEILLFEVDEAIVARKLASEHNQGIQTOAHVSG"
3050..3322
/gene="carBa"
3050..3322
/gene="carBa"
/codon_start=1
/transl_table=11
/product="subunit of meta-cleavage enzyme"
/protein_id="BAA21730.1"
/db_xref="GI:2317680"
/translation="MARVEVDRLIQDMSKEGLIGRVIDTPSDVFEVYGLTPPERTAL
LEGTPQALASIGVHPILQMHVLMYKNPEMATHVSIKDYSDMLKGA"
3322..4131
/gene="carBb"
3322..4131
/gene="carBb"
/codon_start=1
/transl_table=11
/product="catalytic subunit of meta-cleavage enzyme"
/protein_id="BAA21731.1"
/db_xref="GI:2317681"
/translation="MGKIVAAGTSHILMSPKGCESAAARVVGIAELGRRLKEARDP
VLVIITSDHMFNINLSMOPRFVVGIAISYTPMGDMIDPRDLVPGSREVGRALQADE
DGFDCQAEYSLDHGINIPILFMGMEIPVPIVNIINDPIPSARRCVALAESIRQ
AIEKRTPDGCRVAVVAGGLSHWLCVPRHGEVSEKFDHVMVDELVRGNAEKLVMGNE
AIIIDOGGNAGVEILLTWIAAVASESSSEKVFYEAMTQWFTGIGMEHFVK"
5014..5337
/gene="carAc"
5014..5337
/gene="carAc"
/codon_start=1
/transl_table=11
/product="ferredoxin component of carbazole
1, 9a-dioxygenase"
/protein_id="BAA21733.1"
/db_xref="GI:2317683"
/translation="MNQIWLKVCVCAASDMQPTIRRVNRVGAAPLAVRVGDQFYATED
TCTHGISLSEGTLDGDIIECPHFGAFNVCTGMPASSPCTVPJGVFEVEVKEGEVYV
AGEKK"
5718..6707
/gene="carAd"
5718..6707
/gene="carAd"
/codon_start=1
```

```

/transl_table=11
/product="ferredoxin reductase component of carbazole
1, 9a-dioxygenase"
/protein_id="BAA21735.1"
/db_xref="GI:2317685"
/translation="MYOLKIEGQAPQCTSGSKSLVALANGIFPFYECASGCGVCK
FELEGNVOSMPDAPGLSRDRKGRHLACQVALSDLRKIVAVQDKYVPTIPISR
MEAEVVEVVALTHDLLSVRLRTDGPANFLPGQFCLVEAQLPGVVRAYSMANLKNPEG
IMEFYIKRVPYTRGRFSPWLPENRKEGARLFLTGPMTGFFRPGTRKSLCIGGAGLSEY
AAAIARASMETDKPKLVFGSRTPRDAVRWIDIDIDEDKLVVQAVTETDLSLWQGP
TGFIHQVWDAALETLPEVEIYLAGEPPMVDATVRMLLKGKVPDRDQIHDAFF"

ORIGIN

Query Match      100.0%; Score 990; DB 1; Length 6881;
Best Local Similarity 100.0%; Pred. No. 1.5e-256;
Matches 990; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  ATGTACCAACTCAAAATTGAAGGGCAAGCGCCAGGACCTGCGGCTCAGGAGAGAGCCTG 60
Db      5718 ATGTACCAACTCAAAATTGAAGGGCAAGCGCCAGGACCTGCGGCTCAGGAGAGAGCCTG 5777

Qy      61  TTGGTCTCAGCACTTGTCTAATCGTATCGGATTTCCGTACGAGTGTGCATCGGAGAGTTGC 120
Db      5778 TTGGTCTCAGCACTTGTCTAATCGTATCGGATTTCCGTACGAGTGTGCATCGGAGAGTTGC 5837

Qy      121  GGAGTATGCAAAATCGAGTTACTCGAAGGGATGTCCAATCAATGTGGCGGATGTCTCCA 180
Db      5838 GGAGTATGCAAAATCGAGTTACTCGAAGGGATGTCCAATCAATGTGGCGGATGTCTCCA 5897

Qy      181  GGACTTTCTTCGCGAGATCGTGAGAGGGCAACCGCCATCTTCGATGCCAGTGCCTGTGCG 240
Db      5898 GGACTTTCTTCGCGAGATCGTGAGAGGGCAACCGCCATCTTCGATGCCAGTGCCTGTGCG 5957

Qy      241  CTCTCAGACTCGCGATCAAGTTCGAGTCCGGCGCTAACTCATGACCTGCTGCTCGG 300
Db      5958 CTCTCAGACTCGCGATCAAGTTCGAGTCCGGCGCTAACTCATGACCTGCTGCTCGG 6017

Qy      301  TCAAGATGGAAGCGCAAGTTGTGAGTCCGGCGCTAACTCATGACCTGCTGCTCGG 360
Db      6018 TCAAGATGGAAGCGCAAGTTGTGAGTCCGGCGCTAACTCATGACCTGCTGCTCGG 6077

Qy      361  CGATTACGCACTGATCGGCGAGCAAAATTTCTCCCGCGGAGTTCTGCTCTAGTAGGCA 420
Db      6078 CGATTACGCACTGATCGGCGAGCAAAATTTCTCCCGCGGAGTTCTGCTCTAGTAGGCA 6137

Qy      421  GAGCAGTTGCGAGCGCTGTTGCGCATATTCAATGCGCAATTTAAGAACCCCGAAGGC 480
Db      6138 GAGCAGTTGCGAGCGCTGTTGCGCATATTCAATGCGCAATTTAAGAACCCCGAAGGC 6197

Qy      481  ATATGGAGTTCTPATTTAAGAGGTACCCACAGGACGATTTAGTCTCTTGGCTTTTCGAA 540
Db      6198 ATATGGAGTTCTPATTTAAGAGGTACCCACAGGACGATTTAGTCTCTTGGCTTTTCGAA 6257

Qy      541  AATAGAAAAGAGCGCTGCTCTATTTTTCGCGGACCAATGGGACATCTTCTTCCGT 600
Db      6258 AATAGAAAAGAGCGCTGCTCTATTTTTCGCGGACCAATGGGACATCTTCTTCCGT 6317

Qy      601  CCAGGACCGGCGAAGAGTCTTTTCATTTGGCGCGGTGCGCGGCTCTCGTATGCGGC 660
Db      6318 CCAGGACCGGCGAAGAGTCTTTTCATTTGGCGCGGTGCGCGGCTCTCGTATGCGGC 6377

Qy      661  GCTATTGCGCGCTCGATGCGGCAACAGACAGCCCGGTAAAGTTGTTCTACGGCTCA 720
Db      6378 GCTATTGCGCGCTCGATGCGGCAACAGACAGCCCGGTAAAGTTGTTCTACGGCTCA 6437

Qy      721  AGAATCTCCGCGCACGCTGTTCCGTGATCGATATCGATCGATGAGGACAAAGCTTTGAG 780
Db      6438 AGAATCTCCGCGCACGCTGTTCCGTGATCGATATCGATCGATGAGGACAAAGCTTTGAG 6497

Qy      781  GTGCTCCAGGAGTTACGGAAGACAGGATAGCTTTTGGCAAGGGCCCACTGGTTTTATT 840
Db      6498 GTGCTCCAGGAGTTACGGAAGACAGGATAGCTTTTGGCAAGGGCCCACTGGTTTTATT 6557
```

QY 841 CATCAGTTCGACGCGAGCGCTGCTTGAACCCCTACCGGAATACGAAATTTATCTTGC 900
Db 6558 CATCAGTTCGACGCGAGCGCTGCTTGAACCCCTACCGGAATACGAAATTTATCTTGC 6617
QY 901 GGTCCACCGCTATGTCGAGCGTACTGTCGCTGCTGCTGCGCAAGGCTGTTCCACGC 960
Db 6618 GGTCCACCGCTATGTCGAGCGTACTGTCGCTGCTGCTGCGCAAGGCTGTTCCACGC 6677
QY 961 GATCAAAATTCATTTGACGCAATTTTCTAA 990
Db 6678 GATCAAAATTCATTTGACGCAATTTTCTAA 6707

RESULT 3
AB047548 44266 bp DNA linear BCT 03-MAY-2002
LOCUS Pseudomonas resinovorans strain CA10 DNA, car gene cluster flanking
region.
ACCESSION AB047548
VERSION AB047548.1 GI:13094152
KEYWORDS
SOURCE Pseudomonas resinovorans
ORGANISM Pseudomonas resinovorans
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 Sato, S.I., Ouchiyaana, N., Kimura, T., Nojiri, H., Yamane, H. and
Omori, T.
TITLE Cloning of genes involved in carbazole degradation of Pseudomonas
sp. strain CA10: nucleotide sequences of genes and characterization
of meta-cleavage enzymes and hydrolase
J. Bacteriol. 179 (15), 4841-4849 (1997)
MEDLINE 97386424
PubMed 9244273
REFERENCE
2 Sato, S.I., Nam, J.W., Kasuga, K., Nojiri, H., Yamane, H. and Omori, T.
TITLE Identification and characterization of genes encoding carbazole
1,9a-dioxygenase in Pseudomonas sp. strain CA10
J. Bacteriol. 179 (15), 4850-4858 (1997)
MEDLINE 97386425
PubMed 9244274
REFERENCE
3 Nojiri, H., Sekiguchi, H., Maeda, K., Urata, M., Nakai, S., Yoshida, T.,
Habe, H. and Omori, T.
TITLE Genetic characterization and evolutionary implications of a car
gene cluster in the carbazole degrader Pseudomonas sp. strain CA10
J. Bacteriol. 183 (12), 3663-3679 (2001)
MEDLINE 21264379
PubMed 11371531
REFERENCE
4 (bases 1 to 44266)
Nojiri, H., Omori, T. and Habe, H.
TITLE Direct Submission
Submitted (21-AUG-2000) Hideaki Nojiri, The University of Tokyo,
Biotechnology Research Center; 1-1-1 Yayoi, Bunkyo-ku, Tokyo
113-8657, Japan (E-mail: anojiri@mail.ecc.u-tokyo.ac.jp,
Tel: 81-3-5841-3064, Fax: 81-3-5841-8030)
FEATURES
source Location/Qualifiers
1. .44266
/organism="Pseudomonas resinovorans"
/mol_type="genomic DNA"
/strain="CA10"
/db_xref="taxon:53412"
complement(214. .1929)
/gene="ORF31"
complement(214. .1929)
/gene="ORF31"
/functions="putative membrane spanning protein"
/codon_start=1
/transl_table=11
/product="putative ABC transporter subunit"
/protein_id="BAB32742.1"
/db_xref="GI:13094153"
/translation="MGVINLAHGAIFYNWGAFRCPTPRPERLVLARRVGGSSARPVYTR

gene

CDS

complement(2085. .3245)

/gene="ORF30"

complement(2085. .3245)

/gene="ORF30"

/function="putative ABC transporter subunit binding
protein"

/codon_start=1

/transl_table=11

/product="putative ABC transporter subunit"

/protein_id="BAB32743.1"

/db_xref="GI:13094154"

/translation="MKKQLVGTLLAPVAVPALRPRAQDPVRIPTTLSTPAGVLGD
SRDFOALIDQGGKLGVPVSLVEDDGLQARAKQIIRMSLDGSLYGVIFSNV
LAAVNTATKDFYISNTTSPSSLAGCKPXYFVASYQNDTHEMAGVANDLGYKK
MVILAPNYQGRDALGKFKTEIYTKNLQDLSVELARSLAPDAIQQFHP
GGAGINFAKQYSSGLNKSIPMVVVFVSMDBERMLAATGDKVAKGNIITLWNPSTNPA
NOAFVAFNDKXKEVPTVAAQTYDARLIGAAKTVTDGNLKDPEALRSALRNVTFS
IRGPAFGKQHAVIDWYLLRVEADANGKLVQVPVQIAXAQVDSFAECL"

complement(3405. .4385)

/gene="tnpA3"

complement(3405. .4385)

/gene="tnpA3"

/note="tnpA3"

/codon_start=1

/transl_table=11

/product="transposase"

/protein_id="BAB32744.1"

/db_xref="GI:13094155"

/translation="WKQMTFADARYAGKRRKXELFLIEMDRVVPWKGLIALIEPHY
PKGGRPAYPLMAMLRVHLIQNVWFGYSDPAMEALYETTLROFAGINLERIDPETT
ILNFRRLLEKHELAAGILAVINGVLGRGLSROGTIVDATLINAPSTKNGDKRDP
EMHCTKGNQVYFCMKAHIGADDESLVHVSVGTAAVADVTDVQKLHGVENVVAD
AGYGVKREPHREPREVWQVARRSTYKLDKRSVLYKAKKIEKAKAQVRAKVEHS
FRVLRQGYTKPRGLKATQVLTLFALSLSLWMAHRLITNAGEVRL"

complement(4543. .5571)

/gene="antC"

complement(4543. .5571)

/gene="antC"

/note="antC"

/codon_start=1

/transl_table=11

/product="reductase component of anthranilate
1,2-dioxygenase"

/protein_id="BAB32745.1"

/db_xref="GI:13094156"

/translation="MSHKVAPSPADGKTLFFPVOSHEILLDAALRNGINIPLDREGV
CGTCQRCESQYQDYDDALSAEDLRQKMLTCCOTRVSSDAFYEDPSSICNAP
GFRUKGIVRAVEQVSPNTALHLDAGDQGLDFLFGQVYARLQIPGTHGORSYSPAN
RPNQNLQFLIRLLPDGMSNYIRERSLIGDEILLERAPGARLYRIVDKPLVAVAG
KPLSAFLGMLDEIAERGCCHPHLYYGVNADLCETQRTAAVARIPIPRLHPGS
ATPARTWHKRGYITEHFDLAEFRDQALDMDYLCGPPPMVESIKQWLDTQALDQRLYY
EKPTOSNT"

complement(5590. .6081)

/gene="antB"

complement(5590. .6081)

/gene="antB"

/note="antB"

/codon_start=1

/transl_table=11

/product="small subunit of terminal oxygenase component of
anthranilate 1,2-dioxygenase"

/protein_id="BAB32746.1"

/db_xref="GI:13094157"

DEFINITION Pseudomonas resinovorans plasmid pCARI, complete sequence.
ACCESSION AB088420
VERSION AB088420.1 GI:26106080
KEYWORDS Pseudomonas resinovorans
SOURCE Pseudomonas resinovorans
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE
1 Maeda, K., Nojiri, H., Shintani, M., Yoshida, T., Habe, H. and Omori, T.
Complete nucleotide sequence of carboxyle/dioxin-degrading plasmid pCARI in Pseudomonas resinovorans strain CAL0 indicates its mosaicity and the presence of large catabolic transposon Tn4676
J. Mol. Biol. 326 (1), 21-33 (2003)
MEDLINE 22435905
PUBMED 12547188

REFERENCE
2 (bases 1 to 199035)
Nojiri, H., Shintani, M. and Maeda, K.
Direct Submission
Submitted (22-JUL-2002) Hideaki Nojiri, Biotechnology Research Center, The University of Tokyo; 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan (E-mail:anojiri@mail.ecc.u-tokyo.ac.jp, Tel:81-3-5841-3064, Fax:81-3-5841-8030)
LOCATION/Qualifiers
1. 199035
/organism="Pseudomonas resinovorans"
/mol_type="genomic DNA"
/db_xref="taxon:53412"
/plasmid="pCARI"
join(185744..199035,1..59548)
/transposon="Tn4676"
complement(join(198367..199035,1..81))
/gene="ORF18"
complement(join(198367..199035,1..81))
/gene="ORF188"
/function="putative ABC-transporter subunit"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative ABC-transporter subunit"
/protein_id="BAC41710.1"
/db_xref="GI:26106270"

repeat_region
gene
CDS
/translation="MNPQLSVQALEKTFGAVKATNGASFDAFAGELHAIIGNGAGKS
TLTAQAGEPRKGRIFLNGQDITVPAHQRLGLARAFQVSLYPTFLLENVAV
AIARQDQTLGKWPASDQSLAPADYLLARVGLRGAHQVSLSHGRRGLEIAL
ALAEASVLLIDEPMAQAGSARMTQLLELKGRYAILLVEHDMDAVFAALADRTY
LVVGTLLTGSQVRNDPHVRAAYLGEHA"
/gene="ORF31"
complement(1108..1929)
/gene="ORF31"
/function="putative membrane spanning subunit"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative ABC transporter subunit"
/protein_id="BAC41521.1"
/db_xref="GI:26106081"

translation="MGVINLAHGAFTVMGAFCAAYTASLSGSFWLAGLAIVGSALYGL
LLIETSMRLYNRDLQVLAIVVFTNELITVLFGRSPAMPPTPEWYSFVEVL
PGLMYPVSRLLFTAGAVVALAMWLLNHLRLGMLIPAGADHMGVAGLVNTRLYT
LVFVGVCLGLAGVAPAPLLSVIEIGKEKVLITTFVIVVGVGVGRGLAGALLIG
MVDGIGRAYIPQLLDQLDPAEVSGLTSGGLISASAYIVMAVILLVPRGLLPARA"
/gene="ORF30"
complement(2084..3244)
/gene="ORF30"
complement(2084..3244)
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="putative ABC transporter subunit"
/protein_id="BAC41522.1"
/db_xref="GI:26106082"

/translation="MKKQLVGTLLAASCLASLAQAQDPVIRIGFITLSTPAGVIGED
SRGFOAIADQEGKLGVPVSLVDDGLQARAKQIDRMSLQGLSLVYGVIPSNV
LAAVNTATKDGFIYISNTGSPSLAGKQCPHYFVASYQNDTHEMAGVAANDLGYKK
MVLAPNYQGRDALSGFKRIFKEGVEIYIKNLQDLFSVELARKVSLADALFQFHP
GGAGINFAKQYSSGLNKSIPMVVFVSDMERMLAATGDAKGNVILVWNPSTNPA
NQAFVKANDKYKVPVYAAQTYDARLIGAAALKTVGNLKOPEALRSALRNVTFS
IRGFAFGKQHAVIDWYLLRVEADANGKLVQVPVQVITAKAQVDSFAEACAL"
/3255..3258
/rpt_type=direct
3259..4448
/insertion_seq="ISPre2"
3259..3274
/rpt_type=inverted
complement(3404..4384)
/gene="tnpA3"
complement(3404..4384)
/gene="tnpA3"
/notes="transposase of ISPre2"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC41523.1"
/db_xref="GI:26106083"

/translation="MKOMTADAEYAGKQKELFLIEMDRVVPWKGLIALTEPHY
PKGGGRPAVFLMAMLRVHLLQNFYGYSDPAMEALYETTLIQFACGLNLERIPDET
ILNFRLLKHELAAAGILAVINGYLGRLSRLQGTIVDAILINAPSTKXKDKRDP
EMHQTKNQYIFGMAHIGADDESGLVHSVGAANVADVQDKLHGDENVVCAD
AGYTKVKEGHEGRIWQVAARSTYKLDKESVLYKARKIEKAKQVRAKVEHP
FRVIRQFGYTKVFRGLAKNTAQLVTLFALSNNLMARRHLLTNAGEVRL"
/4433..4448
/rpt_type=inverted
4444..4447
/rpt_type=direct
complement(4543..5568)
/gene="antC"
complement(4543..5568)
/gene="antC"
/notes="AntC"
/codon_start=1
/transl_table=11
/product="reductase component of anthranilate"
/protein_id="BAC41524.1"
/db_xref="GI:26106084"

/translation="MSHKVAFSADGKTLFFPVQSHIELLDAALRNGINIFLDCREGV
GTCQRCESQSYVDYDDEALRQRMKLTCTQTRVSSDAAFDFDASSLCNAP
GPGKLGTVRAEQVSNATLHLDAGADGQDLPLQVYARLQIPGTHGORSYFAN
RPSGNOLQFLIRLLPGLMSNVIRESLIGDELLLEAPLGAFLRVHDKFLVFAVG
TGLSAPFLMLDEIAERGCCHPHVLYGVVRNAADLCETQRIAAVAERIPGFRFTPVIS
DPSDWHGKRGYITHEFDLAEFRQALDMLYLCGPPPWVESIKQWLDLQALQDTRLYE
KFTQSNIT"
complement(5587..6078)
/gene="antB"
complement(5587..6078)
/gene="antB"
/notes="AntB"
/codon_start=1
/transl_table=11
/product="small subunit of terminal oxygenase component of
anthranilate 1,2-dioxygenase"
/protein_id="BAC41525.1"
/db_xref="GI:26106085"

/translation="MNSLOQYRVEQFLYRNAEACDQMDTYLSVFSDDSEHFLPQWD
SEHETROPKAMSLIYFNRGGLEDNRVFRIRTKSASTVMPRTILHLLSNRIAPLA
DDELEQVNWHTLYRLQTAEQFFGRATYRURPDGSEWKIARKVILLINDTINSVLDF
YHL"
complement(6081..7484)
/gene="antA"
complement(6081..7484)
/gene="antA"
/notes="AntA"
/codon_start=1
/transl_table=11
/product="large subunit of terminal oxygenase component of

Query Match	100.0%;	Score 990;	DB 1;	Length 199035;
Best Local Similarity	100.0%;	Pred. No. 1.7e-256;		
Matches 990;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1	ATGTACCAACTCAAAATTGAAGGCGAAGCCAGGACCTGCGGCTCAGGGAGAGCCTG	50		
33633	ATGTACCAACTCAAAATTGAAGGCGAAGCCAGGACCTGCGGCTCAGGGAGAGCCTG	33692		
61	TTGTCTTCAGCACTGTGTAATGTGTATCGAATTCGTACGAGTGTGCATTCGGAGGTTGC	120		
33693	TTGTCTTCAGCACTGTGTAATGTGTATCGAATTCGTACGAGTGTGCATTCGGAGGTTGC	33752		
121	GGAGTATGCAAAATTCGAGTTTACTCGAAGGGAATGTCCAAATCAATGTGGCCGGATCCTCA	180		
33753	GGAGTATGCAAAATTCGAGTTTACTCGAAGGGAATGTCCAAATCAATGTGGCCGGATCCTCA	33812		
181	GGACTTTCTTCGCGAGATCTGTAGAAGGGCAACCGGCATCTTTCATGCCAGTTCGCTTGGC	240		
33813	GGACTTTCTTCGCGAGATCTGTAGAAGGGCAACCGGCATCTTTCATGCCAGTTCGCTTGGC	33872		

REFERENCE	2 (bases 1 to 6880)	
	Ouchiyaama,N.	
AUTHORS	Direct Submission	
	Submitted (10-MAR-1997) Naoki Ouchiyaama, Chemicals Inspection &	
JOURNAL	Testing Institute, Kurume Research Laboratories; Chuo-machi 19-14,	
	Kurume, Fukuoka 830, Japan (E-mail:KYS04304@niftyserve.or.jp,	
FEATURES	Tel:0942-34-1500, Fax:0942-39-6804)	
	Location/Qualifiers	
source	1. .6880	
	/organism="Pseudomonas stutzeri"	
CDS	/mol_type="genomic DNA"	
	/strain="OM1"	
CDS	/db_xref="taxon:316"	
	/clone="PSB185"	
CDS	571..1725	
	/codon_start=1	
CDS	/transl_table=11	
	/product="carbazole dioxygenase subunit"	
CDS	/protein_id="BAA31266.1"	
	/db_xref="GI:3293058"	
CDS	/translation="MANVDAIILKVKWAPYDAKLGFRNHYPVWFSKEIDEGEPK	
	TLKLGNIILNVRIDGKLYCKDRLHGVQVSKVECKTKSTITCWIHWIYRWEDG	
CDS	VLCDLINPTSAIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPNFLLDDMEI	
	LGNQIILKSNRLAVENFPDPISHYIHKDSILVKNDLALPLGFAPGGRKQOTRVVD	
CDS	DDVVRGVLDLIGEHGVPVEGTIGEVREGAEGEKIVANDISLWPLGVKVPFP	
	NPDMPQEWVPIDENHYFYFTLGRKCANDEKKNYQEFESKWKPMALGEGFNDDI	
CDS	WAREAMVDYFADDKGWNEILFEVDEALVAWRKLAHEHNGIQTQAHVSG"	
	1837..2991	
CDS	/codon_start=1	
	/transl_table=11	
CDS	/product="carbazole dioxygenase subunit"	
	/protein_id="BAA31267.1"	
CDS	/db_xref="GI:3293059"	
	/translation="MANVDAIILKVKWAPYDAKLGFRNHYPVWFSKEIDEGEPK	
CDS	TLKLGNIILNVRIDGKLYCKDRLHGVQVSKVECKTKSTITCWIHWIYRWEDG	
	VLCDLINPTSAIGROKLYTPVQEAAGCVFIYLGDDPPPLARDTPNFLLDDMEI	
CDS	LGNQIILKSNRLAVENFPDPISHYIHKDSILVKNDLALPLGFAPGGRKQOTRVVD	
	DDVVRGVLDLIGEHGVPVEGTIGEVREGAEGEKIVANDISLWPLGVKVPFP	
CDS	NPDMPQEWVPIDENHYFYFTLGRKCANDEKKNYQEFESKWKPMALGEGFNDDI	
	WAREAMVDYFADDKGWNEILFEVDEALVAWRKLAHEHNGIQTQAHVSG"	
CDS	3049..3321	
	/codon_start=1	
CDS	/transl_table=11	
	/product="2'-aminobiphenyl-2,3-diol 1,2-dioxygenase"	
CDS	/protein_id="BAA31268.1"	
	/db_xref="GI:3293060"	
CDS	/translation="MARYEVDRLIQMSKKEGLIGRIDTSPDVFEYGLTPPERTAL	
	LEGTPQALASIGVHPILQMHYLMYKNPEMATHYSIKDYSMLKGG"	
CDS	3321..4130	
	/codon_start=1	
CDS	/transl_table=11	
	/product="2'-aminobiphenyl-2,3-diol 1,2-dioxygenase"	
CDS	/protein_id="BAA31269.1"	
	/db_xref="GI:3293061"	
CDS	/translation="MGKIVAAGTSHILMSPKCESAARVNVNGIAELGRRILKEARPD	
	VLVITSDHFNINLSMQPFFVVIADSVTPMGDMIDPDLVPGSREVGRAIQAQDE	
CDS	DGFDLQAEYSIDHGMIPILPMGKEIPVFPVIVINTDIPSNARCVALAESIRQ	
	AIEKRPDGRVAVAGAGLSHLCVFRGEVSEKFDHVMDELARKNABELVAMGNE	
CDS	AIIQQGNAGVEILTITWMAVAASESSGEKVFYEAMTQWFTGIGGMEFHVK"	
	4120..4992	
CDS	/codon_start=1	
	/transl_table=11	
CDS	/product="2-hydroxy-6-oxo-6-(2'-aminophenyl)hexa-2,	
	4-dienate hydrolase"	
CDS	/protein_id="BAA31270.1"	
	/db_xref="GI:3293062"	
CDS	/translation="MLNKAQISEKSESAYVERFVNAVGVETRYLEAGKQGPVILIHG	
	GGAGSEGVNRNRPILARHVRVIMLMDLFGKTAQPIETQRRIRHLHDFTKAM	
CDS	NPDGKYSIVGNSMGATGLGVSVLHSELVNLVMSAGLVVEIHEDLRPIINYDPT	
	EGMVLVKTLDNGYKIDDAINSRYTATDEATKAYVATWQWIREOGGLEVDPEFI	
CDS	RKVPVPTLVVHGKDKVVPVETAYKFLDLIDDSWGYIIPHCHWAMIERPDPANAL	
	SFLSRADITRAAA"	
CDS	5013..5336	
	/codon_start=1	
CDS	/transl_table=11	
	/product="ferredoxin reductase"	
CDS	/protein_id="BAA31271.1"	
	/db_xref="GI:3293063"	
CDS	/translation="MNOIWLKVCASADMQPGTIRRVNRVGAAPLAVRVGQDFYATD	
	TCTGIASLSEGTLDGVDIECPFHGAFNVCTGMPASSPCTVPLGVEVEKEGEVYV	
CDS	AVEKK"	
	5370..5687	
CDS	/codon_start=1	
	/transl_table=11	
CDS	/product="unknown"	
	/protein_id="BAA31272.1"	
CDS	/db_xref="GI:3293064"	
	/translation="MADLSVITERVTKAVGNSGLDAVVKDFEPEGVIHIDGMSIPN	
CDS	RVSNEDLPSTITIKILENFEKILNQDLGPKVALATGRMLRGDIRIATRLDKVFLGA	
	PSM"	
CDS	5717..6706	
	/codon_start=1	
CDS	/transl_table=11	
	/product="ferredoxin reductase"	
CDS	/protein_id="BAA31273.1"	
	/db_xref="GI:3293065"	
CDS	/translation="WYOLKIRGAPGTGSGSKLLVSALANGIGFPYECASGGCGVCK	
	FELLEGVQSNWPDAPGLSREKGRHLACQVALSDLRIKVAQDKVPTPIR	
CDS	MEAEVVEVRLTHDLSVRLTDGPNALFQFCVLEAEQPLGVVRAVYANLKNPEG	
	IWEYIRKVPTRFSPMLFENRKEGARLFTGPMGTFFRPGTGRKSLCIGGGAGLSY	
CDS	AAAIARASMETDKFVXLFGYRTPRDARWIDIDEDKLEVVQVMEYEDTSLWQGP	
	TGFIHQVVDAALETLPYEYILAGPPPMVDATVRMLLGKVPREDQIHFDAPF"	

gene /protein_id="BAC56734.1"
/db_xref="GI:28201199"
/translation="MYGPAWYKVDNVRNHTIALILVATLACNBSPLGCTWK
ADGLVPMWTFPRGTEAMGVIEKVYTTNGNVKVTYKGLMKSAIAFTIVNRNTA
NNMYTLHRIR"
5647..6171
/gene="ORFU7"
CDS /gene="ORFU7"
5647..6171
/function="putative cold shock protein"
/codon_start=1
/transl_table=11
/product="cold shock protein"
/protein_id="BAC56735.1"
/db_xref="GI:28201200"
/translation="MELQGRVIEWDERGFIIQNGTSEFTLHISALRKTGRPLL
GSLVTEIRDDGRQLQVAALIGTKSLRQSSGGFSSAVLGLVLLMFGVGVYRF
SNOGTSISASVYKVLARALRPHPEFQCTPEKSKSCASCAEALFHQRCNVKMD
GDGDIPEBQWCN"
6306..6818
/gene="ORFU8"
CDS /gene="ORFU8"
6306..6818
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56736.1"
/db_xref="GI:28201201"
/translation="MKIFASSLAILLATAVPSTAQVATPASEARETLRLADGSRWM
SGDVRGQVYTSMAAALQGTFTENRMKDVQVPAQVESVVFIFGPKASQTVIAHMD
NHGAKYSIPHGTCQITGNSIQLLPYQSGNFRNTMYHGDTSSTVFLESAQPDGSK
HFAEYEVKRN"
6950..7378
/gene="ORFU9"
CDS /gene="ORFU9"
6950..7378
/function="putative transposase ORFA protein"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC56737.1"
/db_xref="GI:28201202"
/translation="MARALLPEDLWLSLIAHLFAHRRSPKGRPRIDRAAPTGLFV
LKTGIPWEYIPREIGCGSGMTWRLHEWMLAGVWQRIHETILRLRHEHQINWDRAS
VDAASVAPAGGHTGRNPTDRGKLGCKHLWLSVGFHW"
7348..7629
/gene="ORFU10"
CDS /gene="ORFU10"
7348..7629
/function="putative transposase"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAC56738.1"
/db_xref="GI:28201203"
/translation="MVDERGFPLVAQISGAQVHDSRFLPLVESVPAVKLAGARKR
FGKUHADRAYSRAHRAWRARRGTAARIARYGVESRRLGRWRWVVERT"
complement(7862..8617)
/gene="car8"
CDS /gene="car8"
complement(7862..8617)
/function="putative transcriptional regulator of car
operon"
/codon_start=1
/transl_table=11
/product="regulatory protein"
/protein_id="BAC56739.1"
/db_xref="GI:28201204"
/translation="MDNFSNLDQAPKASARAVKETSHEHVERQSRQVVEIAKA
IANGLLGRFAPGRLVEADLTKEGLSRNPLREAFGRLEAGLLQVEPRGAIVARL
SRKELVDIFSRECLSEWAARLAENIDKSPFRNYLEASLRNPSTLIPNVSDYLE
DNRKPHSLILKLAGNSHLPKLLDQFPFIQSAPFRMYDANIYQNSTAEHYEILQAIL

gene AGDADAASTRVAARVHRTSKLYSLPDHFFNSD"
complement(8620..8880)
/gene="ORFU12"
CDS complement(8620..8880)
/gene="ORFU12"
/function="unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC56740.1"
/db_xref="GI:28201205"
/translation="MILQVFLDNLNAINAFCDIGLSPFEQAASVCYFRKIALQNC
SNTSFNFRYIQNFLLIKFLKDWNRMTVSNAGHLVKNKLGK"
9141..10130
gene
Query Match 88.0%; Score 871.6; DB 1; Length 20017;
Best Local Similarity 92.5%; Pred. No. 1.6e-224;
Matches 916; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 1 ATGTACCACTCAAAATTGAAGGCAAGCGCCAGGAGCTCGCGCTCAGGGAAGAGCGCTG 60
DB 14091 ATGTACCAACTCAAAATTGAAGGCAAGCGCCAGGAGCTCGCGCTCAGATAAGAGTTTA 14150
QY 61 TTGGTCTCAGCACTTGCTTAATGGTATCGGATTTCCGTACAGTGTGCATCGGAGGTTGC 120
DB 14151 TTGGTCTCGGCGCTCGCAACGGTATCGGACTTCGGTACAGTGTGCTCGGAGGTTGC 14210
QY 121 GGAGTATCAAAATTCGAGTTACTCGAAGGGAATGTCCTCAATCAATGTGGCCGAGTCTCCA 180
DB 14211 GGAGTATGTAATTCGAGTTACTTGAAGGGAACCGTCCCAATCAATGTGGCCGAGTCTCCA 14270
QY 181 GGACTTTCTTCGCGAGATCGTGAGAAGGGCAACCGCCATCTTGCATGCCAGTGGGTTGCG 240
DB 14271 GGACTTTTCATCTCGAGATCGTGAGAAGGGCAACCGTCCACTTGCATGCCAGTGTATTGG 14330
QY 241 CTCTCAGACCTGCGGATCAAAAGTCGAGTCGAGGACGAAGTACGTCGCCAACGATTCCAATC 300
DB 14331 CTCTCAGACCTGCGGATCAAAAGTTCGAGTCGAGGACGAAGTATATATACCGAGGATTCCAATC 14390
QY 301 TCAAGATGGAAGCGGAAGTTGTTGAGTCCGGCGCTAACTCATGACTGCTGTCGCTG 360
DB 14391 TCGAATAATGGAGCGGAAGTTGTTGCGTCCGGGCGCTAACTCATGACTGCTATCTGTA 14450
QY 361 CGATTACCACTGATGGCCAGCAAAATTTCTCCCGGCCAGTTCCTGCTAGTAGAGGCA 420
DB 14451 AAATTACGCACTGATGTCAGCAAAATTTCTCCCGGCCAGTTCCTGCTAAATTGAGGCA 14510
QY 421 GAGCAGTTGCCAGGCGTGGTTCCGCGCATATTCATGCGCAATTTAAAGAACCCCGAAGGC 480
DB 14511 GAGCAGTTGCCAGGCGTGGTTCCGCGCATATTCATGCGCAATTTCAATGACCCCGATGCT 14570
QY 481 ATATGGGAGTTCTATATTAAGAGGTTACCCACAGAGGATTTAGTCTTGGCTTTTGA 540
DB 14571 TTCTGGGAGTTCTATATTAAGAGGTTACCCACCGAGGATTCAGTCTTGGCTTTTCGAG 14630
QY 541 AATAGAAAAGAGGCGCTCGTCTATTTTGGAGGACCAATGGGACATCTTCTTCGCT 600
DB 14631 AATAGAAAAGTAGGCGCTCGTCTATTTTGGAGGACCAATGGGACATCTTCTTCGCT 14690
QY 601 CCAGGACCGCGCCGAAAGAGTCTTTGCAATTTGGCGCGGTGGCGGCTCTCGTATCGCGCC 660
DB 14691 CCAGGACCGCGCCGAAAGAGTCTAATGCAATTTGGCGCGGTGGCTGCTCGTATCGCGCC 14750
QY 661 GCTATTGACGCGCTCGATGCGGAAACAGACAGCCGTTAAAGTCTTCTTACGCTCA 720
DB 14751 GCTATTGACGCGCTCGATGCGGAAACAGACAGCCGTTAAAGTCTTCTTACGCTCA 14810
QY 721 AGAATCTCCGCGGACGCTGTTTCGGTGGATTCGATATCGACATCGATGAGGACAAAGCTTGAG 780
DB 14811 AGAATCTCCGCGGACGCTGTTTCGGTGGATTCGATATCGACATCGATGAGGATAAGCTTGAG 14870
QY 781 GTGCTCCAGGAGTTCGGAAGACACCGGATAGCTTTTGGCAAGGCCGCCACTGCTTTATT 840

```

Db      14871  GTCTGCGAGGAGTTACGGAAGACACACGATAGCCTTTGGCAGGCGGCCCATCGGTTTATC 14930
Oy      841   CATCAGGTGTCGACGCGAGCGCTGCTTGAACCTACCGCAATACGAAATTTATCTGCG 900
Db      14931  CATCAAGTTGTGACGCGAGCGCTGCTCGAAACCTACCGGAATACGAAATTTATCTGCG 14990
Oy      901   GGTCACCGGCTATGTCGACGCTACTGTCGGTATGCTCGGCAAGGCGTGTTCACGC 960
Db      14991  GGTCACCGGCGATGTCGACGCGCACTGTCGGTATGCTCGGCAAGGCGTGTTCACGC 15050
Oy      961   GATCAAAATTCATTTGACGCAATTTTCTAA 990
Db      15051  GATCAAAATTCATTTGACGCAATTTTCTAA 15080

RESULT 7
AF065891
LOCUS      AF065891      8620 bp      DNA      linear      BCT 14-NOV-1998
DEFINITION Ralstonia eutropha JMP134 phenol hydroxylase gene cluster, complete
sequence.
ACCESSION AF065891
VERSION   AF065891.1  GI:3172505
KEYWORDS
SOURCE    Ralstonia eutropha
ORGANISM  Ralstonia eutropha
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Ralstonia.
REFERENCE 1 (bases 1 to 8620)
AUTHORS  Ayoubi, P.J. and Harker, A.R.
TITLE    Whole-cell kinetics of trichloroethylene degradation by phenol
hydroxylase in a Ralstonia eutropha JMP134 derivative
JOURNAL  Appl. Environ. Microbiol. 64 (11), 4353-4356 (1998)
MEDLINE  95013635
PubMed  9797289
REFERENCE 2 (bases 1 to 8620)
AUTHORS  Harker, A.R. and Ayoubi, P.J.
TITLE    Direct Submission
JOURNAL  Submitted (15-MAY-1998) Microbiology, Brigham Young University,
P.O. Box 25108, Provo, UT 84602-5108, USA
FEATURES
     source
         location/qualifiers
             1..8620
                 /organism="Ralstonia eutropha"
                 /mol_type="genomic DNA"
                 /strain="JMP134"
                 /db_xref="taxon:510"
                 /chromosome="I"
             18..30
                 repeat_region
                     /note="similar to TbuT, XylR, and DmpR binding sites"
                     /rpt_type=inverted
                     49..61
                         repeat_region
                             /note="similar to TbuT, XylR, and DmpR binding sites"
                             /rpt_type=inverted
                             56..68
                                 misc_binding
                                     /bound_moiety="IHP"
                                     /evidence=not_experimental
                                     191..218
                                         promoter
                                             /note="consensus sigma-54 promoter region"
                                             284..1801
                                                 gene
                                                     /gene="phlK"
                                                     284..289
                                                         RBS
                                                             /gene="phlK"
                                                             296..1801
                                                                 CDS
                                                                     /gene="phlK"
                                                                     /codon_start=1
                                                                     /transl_table=11
                                                                     /product="putative hydroxylase component"
                                                                     /protein_id="AAC77380.1"
                                                                     /db_xref="GI:3172508"
                                                                     translation="MALLERAANYDIARTNWTFTYVAESDLFFEVNAGQGVPMIWM
                                                                     EYDEPKTSYEVYRIQREKAGAYSVKAALERSRMFADPGWLSILKHYGAIAL
                                                                     GEYAMGAERMAFGAPGMNRNATFGMLDENHGOLQYFPEDYCPKQRPDWAHK
                                                                     AYHTNQGAIARSTFDLPMRSALBIAVLTFAPETGFTNMQFLGLAADAAGDF
                                                                     TFASLISSIQTESRHAQIGGALQILIANGRKEQAQQLVDVAIARAWRLFSLLTCTS

```

gene

RBS

CDS

gene

RBS

CDS

gene

RBS

CDS

gene

RBS

CDS

gene

RBS

CDS

```

MDYATPLQHRKSEKFEFMTIEWIGVFERTLIDLGLPWYWDWDMINEFYOHHAYQMG
IWFVRPTVWNPAAQMTPCDRDWLEEKYPGNWDTFGKAWDVIIONLLAGRELTVPET
LPIVTCVSPICAVPGNKNVRDYPLEYNGRTVHENSEIDRWVQDDPVYRDLTL
VDRELGHQIPDPDGGALRYNNLAPGEIGDDAHQYAWVEAYRROEQNKAA"
1817..2086
/gene="phlL"
1817..1821
/gene="phlL"
1826..2086
/gene="phlL"
/codon_start=1
/transl_table=11
/product="putative hydroxylase component"
/protein_id="AAC77381.1"
/db_xref="GI:3172507"
/translation="MALFPVISNFQYDFVLQVAVDTENSMDEVAASAAAHSVGRRA
POPGKVVRRSGGQGFYPRDARIGDTDIKPMESLEFFCDA"
2085..2429
/gene="phlM"
2085..2089
/gene="phlM"
2094..2429
/gene="phlM"
/codon_start=1
/transl_table=11
/product="putative ferredoxin"
/protein_id="AAC77382.1"
/db_xref="GI:3172508"
/translation="MNFQKSVTLDELWEGDMAEYVDGHVILVRPEGAPRAFQIC
PHODIPLAEGKFGDGRVLMCRHQWTFDANTCKGINPGGCLAEYAVKVDGDDILIAVE
GVEPLFANC"
2438..2765
/gene="phlN"
2438..2441
/gene="phlN"
2449..2765
/gene="phlN"
/codon_start=1
/transl_table=11
/product="putative hydroxylase component"
/protein_id="AAC77383.1"
/db_xref="GI:3172509"
/translation="MSKDHNTAEAYRNRRNVGVFLRASSITSGVIEAAQEDNPGKEIRV
DDKLAYRIDTDSLLIRATLEDALGRPFMRSELVNLSFAGRIETTDYVRFYE
KTL"
2769..3777
/gene="phlO"
2769..2773
/gene="phlO"
2779..3777
/gene="phlO"
/codon_start=1
/transl_table=11
/product="putative hydroxylase component"
/protein_id="AAC77384.1"
/db_xref="GI:3172510"
/translation="MTTQDVLKPLKWTSHLAARRKPSYEIVSTNLHYTTDNPDAP
FELDPNFEMAGWFKRHNASPLKADWNAFRDPDELVYRTYNLQDQGOETVSGLLDQ
PSERGHSMLSHWAGTLRLYTPARYLFAHOMGSAYLTOMAPASTISNCAAYOTAD
SLRWLTHTAYTRBELSOTFGDVGCTDERRYEQDPAWQMKLVEHALVADWAECE
VAFSLVLRPAMEEAVLRGLGEAAHNGTLLGLTDAQLADYQVRRHWAGALVEMALE
TPGNRDVLAGSTARAPLADDAISAYCAALPDAPNAKARACAARVDFWDSIGLAGL"
3818..4824
/gene="phlP"
3818..3821
/gene="phlP"
3826..4824
/gene="phlP"
/codon_start=1
/transl_table=11
/product="putative electron transfer component"
/protein_id="AAC77385.1"
/db_xref="GI:3172511"

```



```
/translation="WKYQISLEGGAVFTVAABEDTLRLGALRAGNALPHECSVGGCGA
CAPDLDGLMESVFEAPGLSERDRKRGYLACQRPUSDCTTIRVCDESYPAPVRAH
RAAALLARALTDMSEPTFRVPQATEFRQYALLYPPRAPGARAYSMANLNEEG
IWKVIRVPAGGAGNALFDQVIGDSVLDGPYGHAYLREDSDARDIVCIAGGSLAP
MLSVARGALAGSSRVHFFYGARGQADIGALDALEKRAEDKRVTLVALSAPSTWK
GPTGFVHEVERNLATSLGSDYDFYFAGPPLMIEAMQALLMKHQVFPFQIRPDRFV"
5051..6837
/gene="phlR"
5051..5054
/gene="phlR"
5059..6837
/gene="phlR"
/codon_start=1
/transl_table=1
/product="putative transcriptional activator"
/protein_id="AAC77386.1"
/db_xref="GI:3172512"
/translation="WAKRHPGPKSGSPAQHPGPNQKRESPEAIRVPAIHDLAKRL
FAPQGRWLDQRMWMLHISLALROELIESLKGKTARGLITRIGQAGTRDAAMA
KRVKRALVDDFLAGPQLVLEGMVHCEPVALDIDVORHYFGDFYLVDSSEAEHI
AGYIGNESVCNMLIGYACGYTSAPMGRPIWRETECGMHVKRVVGVKPVSEWEDA
ODDLRFLRIGDFVKAPVDEAALPATSRIAARLSAPENSFGVVGISAGFNTVCHVN
KVAPDATTVLFGESGVKEVPANLHLRLSRADGPYAVNCASIPHELMSESLFQVE
RGFGTATSRAGRERADGGFLFDEVTGLSFTAGKLLRALQGBIERVGDITRKR
VNRVVAATNLRBAVAGHFREDLFLNLPVFPFQVPLRERRDDIPLMNWFLQRM
ARKDHKHTGFRERADVAMFYDNFQNVNRELENMTERAVILAEQDGLDLCHLTSGE
EYDTTAFMLKRSIGSRSEASESPSAAREPGLAETEVAMLRRAVAEANGNLSR
AARVLGISRPTLAYRLQYGITPEAQ"
complement(6890..8309)
/gene="phlX"
complement(6890..8305)
/gene="phlX"
/codon_start=1
/transl_table=1
/product="putative transport facilitator"
/protein_id="AAC77387.1"
/db_xref="GI:3172513"
/translation="WPRASLSPENDKELVRGNYVRVAVALCAACAGTPAGATDVFNL
EFGFVSRAMGTGGAAPDUGPAAVMANPATLGMAGNHLHLGLDMVTIDIKANTAT
GSTARGSNNGNPPFFAPAAVYVRHNDIAGLGIAPAGAGLQYGGSSFLSTNSG
VQTGLDQFSLRLVIRPPFAVAYRPTDKLTGLASLDVAWTSNLGLTLLDVTQIGLTSQ
RVSGSLVPTLLSFLSGGYLGFNSAPVGGADAWGIGKLGITYDITDPTRVGLA
YNAKTRVGLRQGLATVAVTGNAGNIDLSLPONVYRDQFQWPAQFTVGIHQFNERLSA
ADYQVFWSSVMRDINVTFTRGANLIDLSLPONVYRDQFQWPAQFTVGIHQFNERLSA
HYAORAI PNGMLLAVVLPALPTTLLSGATYAFKDDAIDFALSVDLPKTVGNASQPNL
AVPLIVRHSQNALVLAIRKRP"
complement(8306..8309)
/gene="phlX"
RBS
ORIGIN
Query Match 13.2%; Score 130.8; DB 1; Length 8620;
Best Local Similarity 49.7%; Pred. No. 3.8e-24;
Matches 448; Conservative 0; Mismatches 442; Indels 12; Gaps 4;
QY 53 AGAGCCTGTGTGCTCAGCACTTGCTTAATGGTATCGGATTCCTGACGATGTCATCGG 112
DB 3884 ACACGCTGTGCGCGCGCGCTCGCGCGCGCATGCGCTGCGCGCACAGTGCAGCGTGG 3943
QY 113 GAGGTTCGCGAGTATCAATTCGATTCGAGGGAATGTCCTCAATCAATGTCGCGG 172
DB 3944 GCGGCTGCGCGCGTGCCTTCGACCTTGTGACGCGGCTCAAGAACTCGCTGCGCTG 4003
QY 173 ATGCTCAGGACTTTCTTCGAGATCGTGAGAAGGGCAACGCCCATCTTGTGATCCCACT 232
DB 4004 AGCGCGCGCGCTGTCCGAGCGTGACCGCAAGCGTGGCAA---GTACCTTGGTGCCTAGT 4060
QY 233 CGGTTGCGCTCTCAGACTCGGATCAAGTGCAGTGCAGACAGTACGTCCTCCCAAGA 292
DB 4061 CGCGCGCGTTCAGCGACTGCAGATACGTTGCGCTGCGACCAATCGTATCGCCCGCGG 4120
QY 293 TTCCAAATCTCAAGATGGAAGGGAAGTTGTTGAGTCCGCGCGGTCACTCATGACCTGC 352
DB 4121 TCCGCGCGCATGCCGCGCAGCGGAACTCCTTGCACGCGCGCGGTGACCTTGACATGA 4180
```

```
QY 353 TGTCGCTGCGATTCAGCACTGATGGGCCAGCAAAATTTCTCTCCCGCGCAAGTTCTGCTAG 412
DB 4181 GCGAAATTCAGTTTCCCGGTTCCAGGCGCGACCGAGTTCCGGCGCGCCAGTACCGCTGC 4240
QY 413 TAGAGGCAGAGCAGTTCGCCAGGCGTGTTCGCCCATATTTCAATGCGCAATTTTAAAGAAC 472
DB 4241 TCTACCGCGCCCGTGCACCGGCG---GCCCGCCCTATTTCATGCGCACTTCCGCCAACG 4297
QY 473 CCGAAGCATATCGGAGTTCTATATTAAAGAGGTACCCACAGGACAGATTTAGTCTTGGC 532
DB 4298 AGGAAGGCATCTGGAAGTTCTGATCCGCGCGTTCGCCGCGGCGGCTGGCAGCAACGCGC 4357
QY 533 TTTTCGAATATAGAAAGAGGCGCTCGTCTATTTTTCGCGGACCAATGAGGCACATCTT 592
DB 4358 TGTTGACCAAGTTCGGATCGGGACAGCGTCTGTCTCGACGCGCGCTAGCGCCACGCT 4417
QY 593 TCTTCCCTCCAGGACCGGCCGAAAGAGTCTTTTCGATTTGCGCGCGTTCGCGGCTCTCGT 652
DB 4418 ACCTTCGTAGGACAGTGCCTCGCGCATTTCTGTCATCGCGGTGCTCGGCGCTCGCGC 4477
QY 653 ATGCGCGCGTATTGCAAGCGCGCTCGATGCGCGAACAACAGACAGCG---GTAAAGTTGT 709
DB 4478 CAATGCTGTTCGTCGACGCGGAGCACTCGCGGTAGCGGTTTCGCGCGCTCCACTTCT 4537
QY 710 TCTACGCTCAAGAACTCCCGCGCGAGCGTCTTCGTTGGATTCGATTCGACA---TCGATG 766
DB 4538 TTTATGCGCGCGCGGCGAGGCTGACCTCGTGCCTCGACGCGCTTGGAAAAAATTTGCCG 4597
QY 767 AGNACAGCTTGGTTCGTCAGGCGTTCGACGCGAGTACGGAAGACAGCGATAGCTTTGCGAGGCG 826
DB 4598 AAGACAGCGGGTTACGCTGTCGTCGCGCACTGTCGCGACCGGAGAGCACCTTGAAGGCGC 4657
QY 827 CCATCGTCTTATTCATCAGTTTTCGACGCGAGCGTGTCTTGAACCCCTACCGGAATACG 886
DB 4658 CAACGCGGTTTCGTCAGGAGAGTTCGAGCGTAACCTGACAGCTTCCCTCGCGAGTACG 4717
QY 887 AATTTATCTTCGCGTCCACCGCTATGTCGACGCTACTGTCGATGCTCTCTCGGCA 946
DB 4718 ACTTCTATTTTTCGCGCGCGCGCACTCATGATCGAAGCCATGCGAGCGCTCTCTGATGACA 4777
QY 947 AG 948
DB 4778 AG 4779
RESULT 8
D83068
LOCUS
DEFINITION Pseudomonas aeruginosa DNA for benzene monooxygenase oxygenase
subunit, benzene monooxygenase ferredoxin, complete cds.
ACCESSION D83068
VERSION D83068.1 GI:1754622
KEYWORDS benzene monooxygenase oxygenase subunit; benzene monooxygenase
ferredoxin; bmoA; bmoB; bmoC; bmoD1.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (sites)
AUTHORS Kitayama,A., Suzuki,E., Kawakami,Y. and Nagamune,T.
TITLE Gene organization and low regioselectivity in aromatic-ring
hydroxylation of a benzene monooxygenase of Pseudomonas aeruginosa
J1104
JOURNAL J. Ferment. Bioeng. 82, 421-425 (1996)
REFERENCE 2 (bases 1 to 5337)
AUTHORS Kitayama,A., Kawakami,Y. and Nagamune,T.
TITLE Gene organization and low regioselectivity in aromatic-ring
hydroxylation of a benzene monooxygenase of Pseudomonas aeruginosa
J1104
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 5337)
AUTHORS Kitayama,A.
```

TITLE Direct Submission
JOURNAL Submitted (14-JAN-1996) Atsushi Kitayama, The University of Tokyo,
Department of Chemistry and Biotechnology; 7-3-1 Hongo, Bunkyo-ku,
Tokyo 113, Japan (E-mail: kitayama@tansei.cc.u-tokyo.ac.jp,
Tel: 81-3-3812-2111, Fax: 81-3-5684-8402)

FEATURES
Location/Qualifiers
1. 5337
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="J1104"
/db_xref="taxon:287"
313. .317
324. .1826
/gene="bmoA"
324. .1826
/gene="bmoA"
/note="putative"
/transl_table=1
/product="benzene monooxygenase oxygenase subunit"
/protein_id="BAA11761.1"
/db_xref="GI:1754623"
/translation="MAVLRNIDVDVARTNWTREKYVTEDELPPPELGGSDIPMEKX
EAYDEPKYQTYEYKQREKDAVYVKAALERSKMFENADPGWQSVLKLHGAIPS
GEYAASTAEARMRFPKAPGNRNWATLGSMDIEIRHAQLQYFPHERYKDRQFDWAHK
AFDNEWAAIASRHFDDIMWARDAISVIMLTFEGFTGFMQFLGLAADAEEAGDF
TFSSLISSIQDSESRHAQIGGPTLIILENGKEEAOKYDIAFWARLFSVLGTGP
MDYVTPLEHROSKEFMOEWIVQEFERSIHDLGLDKEWYWDIELEOLDQDHGMILG
VYWRPTWNNPTAGVTPERDMLKRYKRWNDTGHCMQVILDNLYEGRTSLTPT
LPVCMNCLINTTPGNWNVQDYSLEYNGRLHFGESEPDWIFQEPERYAGHMLL
VDRFLGIQMDLGGALAYMDLAPSGSDDAHGYSVWEVYKQLRTKKAS"
1894. .1898
1906. .2172
/gene="bmoB"
1906. .2172
/gene="bmoB"
/codon_start=1
/transl_table=1
/protein_id="BAA11762.1"
/db_xref="GI:1754624"
/translation="MALIPLISNFEFGFVLQVLPVDVNTMTDEVAALAAHHSVGRVK
PRPGHIVRVKQSGSEELIPRTAKISEAGIKMECLFWEAPKH"
2214. .2217
2221. .2556
/gene="bmoC"
2221. .2556
/gene="bmoC"
/note="putative"
/codon_start=1
/transl_table=1
/product="benzene monooxygenase ferredoxin"
/protein_id="BAA11763.1"
/db_xref="GI:1754625"
/translation="MAFKKVCSLDDLMEGESFEVDQGVLLIWPPEGGLKSFQIC
PHQDPLVGEKFGDKVLICRAHLWQFDACSGKGINPGDCAQAQPVKIEGDDVFDTE
GIQPLFSHS"
2612. .2616
2626. .3069
/gene="bmoD"
2626. .3069
/gene="bmoD"
/codon_start=1
/transl_table=1
/protein_id="BAA11764.1"
/db_xref="GI:1754626"
/translation="MGLRVWLARLYSGRGTAATAAVMMQGGVASVSKSTHEKKEETAM
TSNTEIGHKLNRRVGPVRAGELADVAEAREDNPKGEIRVDDKRAYRIDTDDM
IIRRTTGIALGRPDMPELEINLSSFAGR:QTNDYVRFYAKRV"

Query Match 11.8%; Score 117.2; DB 1; Length 5337;
Best Local Similarity 52.8%; Pred. No. 1.8e-20;
Matches 325; Conservative 0; Mismatches 283; Indels 8; Gaps 3;

ORIGIN

58 CTGTTGGTCTACGCACTTGCTAATCGTATCGAATTCGGTACGAGTGTGCATCGGGAGGT 117
4283 CTCCTTCGGTCCGCGCTCCAGGCCGCTGTCGGGTTCCTCCGTACGAGTGCATTCGGGGGCG 4342
118 TCGGAGTATGCAAAATTCGAGTTTACTCGAAGGGAATGTCCAATCAATGTGGCGGATGCT 177
4343 TCGGAGTATGCGCTTCGAGTTTGTGGAGGGGAGTGCATGTGTGTG--CCGAGCC 4400
178 CCAGGACTTTTCGCGAGATCGTGAGAGGCAACCGCCATCTTCGCAATGCGAGTGCCTT 237
4401 CCGGCTTTACGACCGCTGACCGAGCC--GCAACCGCTATCTAGCTGTGAGTGTGTC 4457
238 GCGCTCTCAGACTCGGATCAAGTCCGAGTGCAGGCAAGTACGTCCCAACGATTCCTCA 297
4458 GCAAGACCGACTCCGATCAAGTGAAGCGCGGTGAGGAATATCGCCCCAAGGTACAG 4517
298 ATCTCAAGATGGAAGCGGAAGTTGTTGAGGTCCGGCGCTAACTCATGACCTGCTGTCC 357
4518 CCGCGGTATTTGTCGCCCGGTTTGTGAGGCGGTACGGGTAAACGATGACATCTTTGGAG 4577
358 GTGCGATTACGCACTGATGGCGCAGCAATTTCTCCCGCGCAGTTCTGCTTAGTAGAG 417
4578 TTCGGTTTCGTGCGGTGATGGCTCGCGAGTTCTCCCGCGCAGTACCGCTGAT--G 4634
418 GCAGACAGTTGCCAGCGCTGTTTCGCGCATATTCAATGGCGAATTTAAAGAACCCCGAA 477
4635 CAACTCGCTGGCGTGGAGGTTTCGCGCTTACTCGATGTCCAAACGCGCAACCTCGG 4694
478 GGCATATGGAGTTCTATATTAAGAGGTACCCACAGGAGATTTAGTCTCTGGCTTTTC 537
4695 GGAGAGTGCATTTCCAGATCCGCGCGGTGCGCAATGGCAAGGAGACGACGTACCTGGCC 4754
538 GAAATAGAAAAAGAGCGCTGCTCTATTTTTGACGGGACCAATGGGACATCTTTCTTTC 597
4755 GGCATCTCAACACAGGAGTTCCGTCGAGTTGAGGCGCCCTATGGCTTGCATTTCG 4814
598 CGTCCAGGACCGCGCGAAAGAGTCTTTGCATTTGGCGCGGTGCGCGGCTCTCGATAGCG 657
4815 AGAACCGAAGTCCGCGGACATCGTCTCGTGGCGGAGGCTCGGGGCTCGGCGCTGTG 4874
658 GCGCTATTGACGCG 673
4875 GTGTCGATCGCGCTG 4890

RESULT 9
E32813
LOCUS 5337 bp DNA linear PAT 31-JAN-2002
DEFINITION Process for producing m-hydroxyphenylacetic acid.
ACCESSION E32813
VERSION E32813.1 GI:18623943
KEYWORDS JP 2000069974-A/1.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1. (bases 1 to 5337)
AUTHORS Kitayama, A., Nagamune, T. and Shimizu, M.
TITLE Process for producing m-hydroxyphenylacetic acid
JOURNAL Patent: JP 2000069974-A 1 07-MAR-2000;
SUMITOMO CHEM CO LTD
COMMENT OS Pseudomonas aeruginosa J1104
PN JP 2000069974-A/1
PD 07-MAR-2000
PF 01-SEP-1998 JP 1998247139
PR
PI ATSUSHI KITAYAMA, TERUYUKI NAGAMUNE, MASATOSHI SHIMIZU PC
C12N15/09, C12P7/42//C12N1/21, (C12N15/09, C12R1:19), (C12P7/42, PC
C12R1:19),
PC (C12N1/21, C12R1:19), C12N15/00, (C12N15/00, C12R1:19) CC
FH Key Location/Qualifiers
FT CDS (324)..(1826)
FT CDS (3095)..(4123).

gene 2670. .3728
CDS 2670. .3728
/gene="tbc2E"
/function="catabolism of alkyl- and chloroaromatic compounds"
/codon_start=1
/transl_table=11
/product="Tbc2E monooxygenase"
/protein_id="AAG40798.1"
/db_xref="GI:11878092"
/translation="MTQAEVLKPLTWSHLAARRKSEYEIVSNLHYTDNPAP
FEIDPNFEMAKQKRNANSLPHDPNAPRDEDELVYRINMLQOQETVYFGLDQ
FSRKHDAWERTWAGTLAKLYTPRYLFTLQMSAYLTLQAPASTINCAAYQAD
SLRWLTHAYRTKELSTFDFGFDEREWEDPAWQWRKLVEHALVANDWAECE
VALSVVRPAVEARLSGEARHNGDITLLGLTDAQADQRHLRWAGALVMALE
QPNREVITGWLAKWEPLADEAIVAYCSALPEAPAAQAAVREFTAPQPRVKDS
PGDIIQNPWPTTHNRR"
3725. .4729
/gene="tbc2F"
/transl_table=1
/product="Tbc2F monooxygenase"
/protein_id="AAG40799.1"
/db_xref="GI:11878093"
/translation="MKHQITIEGSAFSAVAEDTLRLGALRAGIALPHCEKSVGGGA
CRFDLLSGLVESIWEAPGLSERDKRKEHLACQRPGLDCTIRVCDTIRVPSVPG
RWAALQARSELTPDMSEFTFAVHAPAEFRPGVALLYLPPHAPGAYSMNLNADG
IWOQVIRVPGGAGSNALFDVBEIGOTVLDGPGYHAHLRDDNARDIVICAGGSLAP
MLSVARGALAEQAGQVRVFFYGGRSQPDLAGAAALDDLVDNRLALSIVLSAPGPELG
WQGTGFVHAEVERVLVAPLDRFEFFYPAGFPFMIEAVQALLMHTHRVPFEQMHDFRV
"

ORIGIN
Query Match
Best Local Similarity 11.0%; Score 109; DB 1; Length 4729;
Matches 462; Conservative 0; Mismatches 470; Indels 21; Gaps 5;
Qy 53 AGACCTGTGGTCTCAGCACTTCTGCTAATGATTCGGATTCCTGATGTCATCGG 112
Db 3783 ACACCTGTGCTCGAGGTGCGCTGCGCGCGGCATCGCCCTGCCCATGAATGAGCGTGG 3842
Qy 113 GAGGTTCGGAGTATGAAATTCGAGTTACTCGAAGGGAATGCCAATCAATGTGCGCGG 172
Db 3843 GCGCTGCGGCGCAATGCGCTTCGACCTGCTCTCGGGGCTCGTGAATCCATCTGCGCGG 3902
Qy 173 ATGCTCCAGGACTTCTTCGCGAGATCGTGAGAGGCAACCGCCATCTTCATGCGCAAT 232
Db 3903 AGGCGCTGGGCTGTCGGAACGCGACGCGCAAGCGCGCAA---ACACCTCGCCTGCCAAT 3959
Qy 233 GCGTTGCGCTCTCAGACTCGGATCAAGTCGAGTGCAGGACAGTACGTCCCAAGA 292
Db 3960 CGCGCGCTGGGCGATTGACAGATTCGCGGTGCGCTGCGACGACACTACCGCGCGGTGG 4019
Qy 293 TTCCAATCTCAAGATGGAAGCGGAAGTGTGTGAGGTCCGGGCGCTAACTCATGACCTGC 352
Db 4020 TGTCCCGCGCGCTGGCGCGCGCTGTGTCAGGCGCAGCGGGAGCTACGCGCGACATGA 4079
Qy 353 TGTCCGTGCGATTACGACTATGCGGCGAGCAAAATTCCTCCCGCGCGCTTCGCGTAG 412
Db 4080 CGCAATTCACGTTTCGAGTACACGCGCGCGCTGAGTTTCGCCCGGGCGCTAGCGCGTGC 4139
Qy 413 TAGAGGCGAGACGAGTTCGCGCGCTGTTCCGCGATATTCATGCGCAATTTAAAGAAC 472
Db 4140 TGTATCCGCCCGCAGCGCGCTGGCG---CGCGCTATTCCATGTCCAACTGCCCAACG 4196
Qy 473 CCGAAGGCAATGAGGAGTCTATATTAGAGGGTACCCACAGGAGATTTAGTCTTCGCG 532

Db 4197 CAGACGGCATCTGGCAGTCTTGTGATTCGGCGCGGTTCCTCGGCGGGCGCGCAGCAACGCGC 4256
Qy 533 TTTTCGAAATAGAAAAGAGCGCGCTCGTCTATTATTTTACGCGGACCAATGGGCACATCTT 592
Db 4257 TGTTTCGACCGATCGAGATCGGGCAGACCGTACCTCGACGGGCGCTATGGCCATGCC 4316
Qy 593 TCTTCGCTCCAGGACCGCGCGAAGAGTCTTTGCAATTTGGCGCGCGGTGCGCGGCTCTCGT 652
Db 4317 ACCTGCGGACGACAAACGCGCGCGACATCGTGTGATCGCTGGCGGCTCCGCGCTGCTC 4376
Qy 653 ATGCGCGCGCTATTTCACGCGCGCTCGAT---CGCGGAACACAGACAGCGGCTAAAGTTGT 709
Db 4377 CGATGCTGCTCGCTCGCGCGGAGCGCTCGCGAGAGGGCGCGCAACGCTGCACTTCT 4436
Qy 710 TCTACCGCTCAAGAACTCGCGCGACGCTGTTCGGTGGATCGATATCGACATCGATGAGG 769
Db 4437 TCTACGCGCGCGCTCGCAACCCGACCTCGCGCGCATCGCGCGCTCGATGACCTGCTGG 4496
Qy 770 ACAAGCTTGAGGTCTCTCAGCGCAGTTA-----CGGAACACAGCATAGCCTTTTCGC 820
Db 4497 GCGAACACCGCTCGCGCTCTCGGTGCTGTGTCGCGCGCGCGCGCGAATCGCGCTGCG 4556
Qy 821 AAGGCGCCACTGTTTATTCATCAGTTGTGCGACGCGCTGCTTGAAACCCCTACCGG 880
Db 4557 AGGCGCGCGCTGCTGTCATGCGAGGTGCGAAGCGGTGCTGTGCGCGCTGCGGAC 4616
Qy 881 AATAGAAATTTATCTTCGCGCTCCACCGCTTATGTCGATCGACGCTACTGCTGCTATGCTGC 940
Db 4617 GCTTCGAGTCTATTTCGCGCGCGCGCGCGCATATCGAAGCGCTGCGGCAATTCGTGA 4676
Qy 941 TCGGCA---AGGCTGTTTCCACGCGATCAATTCATTTTTCAGCGCATTTTCTAA 990
Db 4677 TGCACACGATCGGTGCGCGTTTGAACAGATGCAATTCGACCGCTTCGCTCTAA 4729

RESULT 11
PPU04052
LOCUS
DEFINITION
Pseudomonas pickettii PK01 tolune-3-monooxygenase oxygenase
subunit (tbaA1) and (tbaA2) genes, (tbaU) gene, (tbaB) gene,
toluene-3-monooxygenase ferredoxin protein (tbaV) gene, and
toluene-3-monooxygenase oxidoreductase (tbaC) gene, complete cds.
U04052
VERSION
U04052.1 GI:1580790
KEYWORDS
SOURCE
ORGANISM
Ralstonia pickettii
Ralstonia pickettii
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Ralstonia.
REFERENCE
AUTHORS
Byrne, A.M., Kukor, J.J. and Olsen, R.H.
TITLE
Sequence analysis of the gene cluster encoding
toluene-3-monooxygenase from Pseudomonas pickettii PK01
JOURNAL
Gene 154 (1), 65-70 (1995)
MEDLINE
95172404
PUBMED
7867951
REFERENCE
AUTHORS
Byrne, A.M. and Olsen, R.H.
TITLE
Cascade regulation of the toluene-3-monooxygenase operon
(tbaA1-tbaA2) of Burkholderia pickettii PK01: Role of the tbaU
promoter (tbaU1) in the expression of its cognate activator, TbaU
JOURNAL
J. Bacteriol 178 (21) (1996) in press
REFERENCE
AUTHORS
Byrne, A.M.
TITLE
Direct Submission
JOURNAL
Submitted (06-DEC-1993) Armando M. Byrne, Department of
Microbiology & Immunology, University of Michigan Medical School,
5641 Medical Science Building II, 1301 West Catherine Street, Ann
Arbor, MI 48109-0620 USA
COMMENT
On Oct 2, 1996 this sequence version replaced gi:505045.
FEATURES
Location/Qualifiers
1. .4993
/organism="Ralstonia pickettii"

```

/mol_type="genomic DNA"
/strain="PKO1"
/db_xref="taxon:329"
18..62
repeat_region

/note="putative TbuT-binding site"
/citation={2}
/rpt_type=inverted
193..207
repeat_region
215..252
/rpt_type=inverted
243..256
promoter
280..284
RBS
293..1798
gene
/gene="tbaU1"
293..1798
CDS
/gene="tbaU1"
/note="putative"
/codon_start=1
/transl_table=11
/product="toluene-3-monooxygenase oxygenase subunit"
/protein_id="AAB09618.1"
/db_xref="GI:505046"
/translation="MALLERAAWYDIARTNWTSPSYVTESLFPDINTGAGVPMETW
ETYDEPYTSYPEYVIOREKDGAYSVKALERSRPEDADPGMLSILKAHYGAIAL
GEYRAMSAEAMARFGRAPGNRMATFGMDLNHRHGLQLYFFHDYCAKORQFWAHK
AYTHWEIGAIARSFDDLFNSRSADIAIMLTAFETGTMTMOFLGLAADAAEGDF
TFASLSIOTDESHEAQIGGPALQLITASCKEQAOXLDVIA TARAWRLFSLLTGNS
MDAPPLPHRKSESKEFEWTVGVOPERTLDLGDLPTKYWDQMINDEFVOHHAYONG
TWFRPTIWNPAACITDCDRNLBEKYPGWNDIFGRANDVIINLLAKRFELTVPEP
LPICVMQSCLPICAVPGNGWIYKDIPDYIKGRTYHNSEIDRWVFQQDPURDHULTL
VDRLFAGIQIQPNLMGALQYNLAPGECGDAAHYAWVEAYRNORYOKKAA"
1815..1818
RBS
1823..2083
gene
/gene="tbaU1"
1823..2083
CDS
/gene="tbaU1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB09619.1"
/db_xref="GI:505047"
/translation="MALLSFVISNFQYDFVLQIVAVDIENTIDEVA AAAAHHSVCRRVA
POFGKIVVRKGGEQIFRNARLATDTIKMEALEFIFCDA"
2082..2085
RBS
2091..2426
gene
/gene="tbaU2"
2091..2426
CDS
/gene="tbaU2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB09620.1"
/db_xref="GI:505048"
/translation="MNFOKVGSDELWBGDMTEVEVAGHVIVLVPRRGEPRAFGIC
PHODIPLAEGRFDGVRVLCRAHQHTFDANTGKINGPGSRLAEALKVBGGDILLAVE
GVBLPANC"
2435..2439
RBS
2446..2760
gene
/gene="tbaU2"
2446..2760
CDS
/gene="tbaU2"
/note="putative"
/codon_start=1
/transl_table=11
/product="toluene-3-monooxygenase ferredoxin protein"
/protein_id="AAB09621.1"
/db_xref="GI:505049"
/translation="WSINDTAAYNRNVGPVLRASSITAGVIEAAQSDNPQRSIRID
DKLAYRIDTDGELLRLRATLEALGRFPKKSELEVNLSFAGRIETDDIVRYRYEK
TL"
2763..2767
RBS
2773..3762
gene
/gene="tbaU2"

```

QY 593 TCTTCGCTCCAGGACCGCGCAAGAGCTCTTTCGATTTGGCGGCTGCGGGCTCTCGT 652
 Db 4419 ACCTGGCGGACGACAAACCGCGCGACATCGTGTGTGATCGCTGGCGGCTCCGAGTGGCTC 4478
 QY 653 ATGCGGCGCTATTGACGCGCTCGAT---GCGGAAACAGACAAAGCGGTAAGTTGT 709
 Db 4479 CGATGCTCTCGTCCGCGCGGCGCTCGCGCAGGAGGCGCGCAACGCGTGCACTTCT 4538
 QY 710 TCTACGGCTCAAGAACTCCGCGGCGCTGTCGGTGATCGATATCGACATCGATGAGG 769
 Db 4539 TCTACGGCGGCGCTCGCAACCGGACCTCGCGGCGCATGGCGGCGCTGATGACCTGTGG 4598
 QY 770 ACAAGCTTGAGTCTCCAGGACAGTTA-----CGGAAAGACACGATAGCTTTGGC 820
 Db 4599 GCGAACACCGCTCGCGCTCTCGTGTGCTCGCGCCCGCGGCACTCGGCTGGC 4658
 QY 821 AAGGCGCCACTGGTTTATTATCATCAGGTTGTCGACGACGCTGTTGAAACCCCTACGG 880
 Db 4659 AAGGCGCCACTGGTTTCTGTGATCGCGAGGTCGAAACGCGTCTGGTGGCGCGCTGAC 4718
 QY 881 AATCGAAATTTATCTTCGCGGCTCCACCGCTATGCTGACGCTACTGTCGGTATGCTGC 940
 Db 4719 GCTTCGAGTCTATTTCGCGGCGCGCGCCCATGATCGAAGCCGTGCGAGCATGCTGA 4778
 QY 941 TCGGCA---AGGCTGTTCCACGCGATCAAAATTCATTTGACGCAATTTCTAA 990
 Db 4779 TGCACACGATCGGTCGCTTTGAACAGATGCATTTGACCGCTTCGCTAA 4831

RESULT 12
 E39081 5743 bp DNA linear PAT 31-JAN-2002
 LOCUS
 DEFINITION Gene encoding desulfurase.
 E39081
 ACCESSION
 E39081.1 GI:18628914
 VERSION
 JP 2000245477-A/1.
 KEYWORDS
 Burkholderia sp.
 SOURCE
 Burkholderia sp.
 ORGANISM
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 Burkholderiaceae; Burkholderia.
 REFERENCE
 1 (bases 1 to 5743)
 Kato,A. and Yanase,E.
 Gene encoding desulfurase
 Patent: JP 2000245477-A 1 12-SEP-2000;
 JOURNAL
 PETROLEUM ENERGY CENTER
 COMMENT
 OS Burkholderia sp.
 PN JP 2000245477-A/1
 PD 12-SEP-2000
 PF 04-MAR-1999 JP 1999057446
 PR
 PI AKIO KATO, EIJI YANASE
 PC C12N15/09, C10G32/00, C12N9/02, C12N9/90/// (C12N15/09, C12R1:01),
 PC C12N15/00,
 PC (C12N15/00, C12R1:01)
 CC

Key Location/Qualifiers
 FT CDS (78)..(1103)
 FT CDS (2499)..(3329)
 FT CDS (3389)..(4744)
 FT CDS (4777)..(5271)
 FT CDS (5455)..(5742).

FEATURES
 source
 1. 5743
 /organism="Burkholderia sp."
 /mol_type="genomic DNA"
 /db_xref="taxon:36773"

ORIGIN

Query Match 10.9%; Score 107.8; DB 6; Length 5743;
 Best Local Similarity 49.8%; Pred. No. 6.2e-18;
 Matches 330; Conservative 0; Mismatches 327; Indels 6; Gaps 2;

QY 40 TCGGCTCAGGGAAGAGCCTGTTGGTCTCAGCACCTTGCTAAATGGTATCGGATTCGGTAC 99
 Db 135 TGTTCACATGATGAACGCTACTCGCTGGCGGCTCAAGGCTGGCTTGGATTTCCCTAT 194
 QY 100 GAGTGTGCATCGGAGGTTGCGAGATATGCAAAATTCGAGTTACTCGAAGGGAATGTCAA 159
 Db 195 GAATGTCAATCGGATCGTGCAGCAGTTGCCGCTTTTCACTGCTGAGGCGATGTCAAG 254
 QY 160 TCAATGTGCGCGATGCTCCAGGACTTTCTTCGCGAGATCGTGAGAAGGGCAACGCCAT 219
 Db 255 GATTGTGTGATCGAATGCGCGGTTTGAATGCGAAGCTCGTGAATGCGGCA---TGCAC 311
 QY 220 CTTGCAATGCCAGTGTGCGTCTCAGACCTTCGCGATCAAAAGTCGACAGTCAGGACAAG 279
 Db 312 CTGCGCTGCAGACACACAGTAGTACTGTGCAATCAAGCTGCGCTTGAAGCGCGGC 371
 QY 280 TAGCTCCCAACGATTCATCTCAAGATGGAAGCGGAAGTTGTTGAGTTCGCGGCGCTA 339
 Db 372 TATGTCCCAGAGTGATGCCGACGCTCAGCATGCGGCGTTGGTCGATGTGTCGCACCTG 431
 QY 340 ACTCATGACCTGCTGTCGTCGATTCAGCACTGATGGCCAGCAAAATTTCTTCCCGGC 399
 Db 432 ACGCAAGACATGTTTCGAGTTCCAGTTTCGTACGGAGGCTCCAGCTGAGTTTTCCTGGC 491
 QY 400 CAGTTCTGCTAGTAGAGGACAGCAGTTGCCAGGCGTGTTCGCGCATATTCATGCGG 459
 Db 492 CAGTATGCGCTGCT---TCGCTTTCCTGCGTGTGCGCACAGGCGCTACTCGATGAGC 548
 QY 460 AATTTAAAGAACCCGAGGACATATGGGAGTTTCATATTAAGAGGCTACCCACAGGACGA 519
 Db 549 AACCTGCCCAATGGCATTTGCGAATGCGATTCATCGTCAAGCGCAACCGGGTGTTGT 608
 QY 520 TTTAGTCTTGGTTTTTCGAAATAGAAAGAGCGCTCGTCTATTTTGGCGGACCA 579
 Db 609 GGAACCGCGTGCTTTTCGATGTCCTGAAACGAGGTGATGCTATTCAGTTGGAGGGTCT 668
 QY 580 ATGGGCACATCTTCTTCGCTCCAGGACCGGCGCAAGAGTCTTTGCAATTTGCGCGGCT 639
 Db 669 TAGCGGAGGCTTATCTGCGTACCGAACCAGGTAGAGGTGCTCTGCAATGCTGCGGCG 728
 QY 640 GCGGCGCTCTGATGCGGCGCTATTTGCAAGCGCTCGATGCGCGCAACAGACAGCGG 699
 Db 729 TCGGAGCTTTCTCCGATGTTGTCATCTCGCGGCGCAGTGTGCAATCGGCAATGACC 788
 QY 700 GTA 702
 Db 789 GAA 791

RESULT 13
 AB024945

LOCUS
 DEFINITION Alcaligenes faecalis phenanthrene degradative gene cluster (phnAb to phnE genes), complete cds.

ACCESSION
 AB024945.1 GI:4586270
 VERSION
 3.4-dihydroxyphenanthrene dioxygenase; phnC; 2-carboxybenzaldehyde
 dioxygenase; phnI; 1-hydroxy-2-naphthoate dioxygenase; phnG;
 trans-2'-Carboxybenzalpyruvate hydratase-aldolase;
 glutathione-S-transferase; isomerase; dioxygenase small subunit;
 dioxygenase large subunit; dihydrodiol dehydrogenase; ferredoxin
 reductase; ferredoxin.

KEYWORDS
 Alcaligenes faecalis
 Alcaligenes faecalis
 Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 Alcaligenaceae; Alcaligenes.
 1 (sites)

REFERENCE
 AUTHORS Kiyohara,H., Tabata,Y. and Takizawa,N.
 TITLE A phenanthrene degradative gene cluster in Alcaligenes faecalis
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 24448)
 AFK2

AUTHORS
 Takizawa,N., Tabata,Y., Fujimoto,T. and Kiyohara,H.

TITLE Direct Submission
JOURNAL Submitted (14-MAR-1999) Noboru Takizawa, Okayama University of Science, Applied Chemistry; 1-1, Ridai-cho, Okayama 700-0005, Japan (E-mail: takn@bioeng.dac.ous.ac.jp, Tel: 81-86-256-9552, Fax: 81-86-256-9552)

FEATURES
Location/Qualifiers
1..24448
/organism="Alcaligenes faecalis"
/mol_type="genomic DNA"
/strain="AFK2"
/db_xref="taxon:511"
1718..2059
/gene="phnA"
1718..2059
/gene="phnB"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="BAA76320.1"
/db_xref="GI:4586271"
/translation="MTKSNATAGDITLPICAADSLAPQVACDVANLPLAIYNL DGEYLATSNICTHEVALLDGFEGELVCEAHGGSFNRTGEAKSLPCKRLQVFSV VKDGGQVFKI"
2083..3111
/gene="phnA"
2083..3111
/gene="phnA"
/codon_start=1
/transl_table=11
/product="ferredoxin reductase"
/protein_id="BAA76321.1"
/db_xref="GI:4586272"
/translation="MSFQSFQVRIGAGDTFVFCSTDETLAAAKAGLGFPPYEQSG SCSSCFQLEGGVGNWAGNLAECGMLGCSQTPGSDCRILKLPKGVPO WITROHVALVPLTQDMFEPQTEAPAEPLPGQYALLRLPGVGTAYSMNSPL NGGEVHEIVKRGCGGTAVLDLKRGALELEGPYGAALVRETGDVVICGGS GLSPMLSLGAVCNPAINTERLLMFIYGTDLHCADVAFGEPLKRVLEFSAIS DVNAESANDWGERGLIHEVLAKHIGPNQGYDFYFCPPMTDAVHKLLVLRYPSE QLHFRFV"
3210..4043
/gene="phnB"
3210..4043
/gene="phnB"
/codon_start=1
/transl_table=11
/product="dihydrodicol dehydrogenase"
/protein_id="BAA76322.1"
/db_xref="GI:4586273"
/translation="MAWLEQSVLTTGGVAGLGRALVKRLVEGANVTVLDRNARGLD ELVESKGRVAGSGDVRLNADNRKAVELAVPEGLDTEFNRQHLGLCPQCRPAG RCQRSPDEVGINLNGYNGIKAAPALVRSGLVILILSSAFYAGCGGVLYTVAK HAAVLGIQAAHELAPYRVNGVAPGSIASDLRGPXSLGNGEQSITTSVPLADLVKDIA PIGRLSDTESYTGSYVYLASARNAPATGVIINCDDGNGVRSVLPASGGKGLLEKFG G"
4100..5413
/gene="phnA"
4100..5413
/gene="phnA"
/note="iron-sulfur protein"
/codon_start=1
/transl_table=11
/product="dioxxygenase large subunit"
/protein_id="BAA76323.1"
/db_xref="GI:4586274"
/translation="MSTNRITQDVESLVDVKSGIQAKRIFWDQIYDLIEIQIFGKCW QFLTHDSLINPGDFTTGYWAEDEVIVNRQDGSVKAFLNVCTRGRARLVAAEANAR AFSCTYHGWFGLDGALKVYPMEKELSDKSKFGLREIRYVESYRGYYGNFDEA PSLNDYGVKFDYLDIWDINGGIELVPPSRSLNKNWKTPAENFVGVAYHVGWTHL ASLMALEALAAIGNEAPNEGIGIOVTSRFGHGFALMDVAPVAHVANSKAGTLYR EWFASRKPQVKEIKGAVGRIYGHSLNGTIFPNSFLIGTNTFLKAVPRGPIQIEYF WAIVEKMPALKNVAVSGNNSFGSAGMLEADDSDNMTWTONNREGMTREGKLSQ LGLGFLDDPDPFGVIGQSAIGETLIEGFTGPRKRLWAQTGERS"
5487..6005

gene
CDS
/gene="phnAd"
5487..6005
/gene="phnAd"
/note="iron-sulfur protein"
/codon_start=1
/transl_table=11
/product="dioxxygenase small subunit"
/protein_id="BAA76324.1"
/db_xref="GI:4586275"
/translation="MNEILPVRGSESEVVSTLYREARLHAHEYEENMMLHPDQY QAYFQQLRYKRYKLPDKVYAIIDRFDFLDARIQFESGVQMSDPPPIRRLSN IEVFKASSTGNQVFLNVFVRNRRVDETTYSYGREEWVRGEGALQLLKRIVSLDE REILLGNVNFEL"
6161..6751
/gene="phnD"
6161..6751
/gene="phnD"
/note="similar to 2-hydroxychromene-2-carboxylate isomerase"
/codon_start=1
/transl_table=11
/product="isomerase"
/protein_id="BAA76325.1"
/db_xref="GI:4586276"
/translation="MLSFYDFDPSFYSLASIELPIVQRYGYSVSKYKPIDIACAKRA IGNVPSNRDMPVKLTHLSRDQLQWQRYGTPLKFPFSDRLNTGFFYAAAGEARA EYVRAPHLTWGMGOAPSGEPFLRSIASEGMNVDDFMQFTDSVDGANAYKQSIDEGI ARSVFGVPMVVICDEMWMGNDRLDFDVEYLQSRASN"
7714..8325
/gene="gst"
7714..8325
/gene="gst"
/codon_start=1
/transl_table=11
/product="glutathione-S-transferase"
/protein_id="BAA76326.1"
/db_xref="GI:4586277"
/translation="WIKKUYGAPGACSLAHIAHLEAGLTNVPVLDLRTQRTDGRS LADNRSYVYVLEDEQILTEVPVLYIADLPEKGLAPAGALPLIDCRSGSUS LTNFTSRFSCRSRSEEVKASRARVQLRTWDAQLAGRTYLVEGQFTVADVYLF VVAGTKFTGTDVQELTRLGSLYLGRIQAQRPSVEAAMEAEGLVK"
8403..9287
/gene="phnH"
8403..9287
/gene="phnH"
/codon_start=1
/transl_table=11
/product="trans-2'-carboxybenzalpyruvate hydratase-aldolase"
/protein_id="BAA76327.1"
/db_xref="GI:4586278"
/translation="WKIIPSDIRGVGINTPATEDASNNQTOOSVNLSETKAVREL NTAKIDILATGTGEGATLWPEKLVDCVQVQTKRSPPLFAGVTILNTRDTIERA RALVDIGVDGLFVGRPMWLMDDKSIYRYQDITARALPGVPMVAYDNPVAFKGIKSTE AYLELAKIPEVVASKTKGCPALAKMDRAVGEKRIILPLATWALAAQELPELALAAWS GSVCAPTNNMALSQAILAGDWQRAKELSEKCNWGVGVHVCGRFGGLYGLIHPDRSF AFCSSRCKSGSSATALP"
9406..10455
/gene="phnG"
9406..10455
/gene="phnG"
/codon_start=1
/transl_table=11
/product="1-hydroxy-2-naphthoate dioxxygenase"
/protein_id="BAA76328.1"
/db_xref="GI:4586279"
/translation="WTMSDHSKPVTEVDVPELQSKYKDSIVFLNWSVPATSGKQI BEAQITWSTVLPITIOETAKQKRVTHLDRVLLMTNPKRLHVDEAGTGPLMDFQCV MPGRAGARHPMNAALFVQCTGVKSIVDGKOCVMEPGDLIITPGWCHEHNEKG DPITWLDVWMLHFLTGTGFGPQARDVRPOLDAFTAAGIIVPKLDGSDHASRSY SPIFRYPWDAVRALVAPSAASDGSREVRVGNPLTGGPSMDLI DSTLIQLESSEKSTRK YKSSANACFFVEGHVGSQIGETIENSHPDVTMPANLWASHQAKGTARIFQVSTR EYIRLGLTFEAFGE"

gene
CDS
/gene="phnAd"
5487..6005
/gene="phnAd"
/note="iron-sulfur protein"
/codon_start=1
/transl_table=11
/product="dioxxygenase small subunit"
/protein_id="BAA76324.1"
/db_xref="GI:4586275"
/translation="MNEILPVRGSESEVVSTLYREARLHAHEYEENMMLHPDQY QAYFQQLRYKRYKLPDKVYAIIDRFDFLDARIQFESGVQMSDPPPIRRLSN IEVFKASSTGNQVFLNVFVRNRRVDETTYSYGREEWVRGEGALQLLKRIVSLDE REILLGNVNFEL"
6161..6751
/gene="phnD"
6161..6751
/gene="phnD"
/note="similar to 2-hydroxychromene-2-carboxylate isomerase"
/codon_start=1
/transl_table=11
/product="isomerase"
/protein_id="BAA76325.1"
/db_xref="GI:4586276"
/translation="MLSFYDFDPSFYSLASIELPIVQRYGYSVSKYKPIDIACAKRA IGNVPSNRDMPVKLTHLSRDQLQWQRYGTPLKFPFSDRLNTGFFYAAAGEARA EYVRAPHLTWGMGOAPSGEPFLRSIASEGMNVDDFMQFTDSVDGANAYKQSIDEGI ARSVFGVPMVVICDEMWMGNDRLDFDVEYLQSRASN"
7714..8325
/gene="gst"
7714..8325
/gene="gst"
/codon_start=1
/transl_table=11
/product="glutathione-S-transferase"
/protein_id="BAA76326.1"
/db_xref="GI:4586277"
/translation="WIKKUYGAPGACSLAHIAHLEAGLTNVPVLDLRTQRTDGRS LADNRSYVYVLEDEQILTEVPVLYIADLPEKGLAPAGALPLIDCRSGSUS LTNFTSRFSCRSRSEEVKASRARVQLRTWDAQLAGRTYLVEGQFTVADVYLF VVAGTKFTGTDVQELTRLGSLYLGRIQAQRPSVEAAMEAEGLVK"
8403..9287
/gene="phnH"
8403..9287
/gene="phnH"
/codon_start=1
/transl_table=11
/product="trans-2'-carboxybenzalpyruvate hydratase-aldolase"
/protein_id="BAA76327.1"
/db_xref="GI:4586278"
/translation="WKIIPSDIRGVGINTPATEDASNNQTOOSVNLSETKAVREL NTAKIDILATGTGEGATLWPEKLVDCVQVQTKRSPPLFAGVTILNTRDTIERA RALVDIGVDGLFVGRPMWLMDDKSIYRYQDITARALPGVPMVAYDNPVAFKGIKSTE AYLELAKIPEVVASKTKGCPALAKMDRAVGEKRIILPLATWALAAQELPELALAAWS GSVCAPTNNMALSQAILAGDWQRAKELSEKCNWGVGVHVCGRFGGLYGLIHPDRSF AFCSSRCKSGSSATALP"
9406..10455
/gene="phnG"
9406..10455
/gene="phnG"
/codon_start=1
/transl_table=11
/product="1-hydroxy-2-naphthoate dioxxygenase"
/protein_id="BAA76328.1"
/db_xref="GI:4586279"
/translation="WTMSDHSKPVTEVDVPELQSKYKDSIVFLNWSVPATSGKQI BEAQITWSTVLPITIOETAKQKRVTHLDRVLLMTNPKRLHVDEAGTGPLMDFQCV MPGRAGARHPMNAALFVQCTGVKSIVDGKOCVMEPGDLIITPGWCHEHNEKG DPITWLDVWMLHFLTGTGFGPQARDVRPOLDAFTAAGIIVPKLDGSDHASRSY SPIFRYPWDAVRALVAPSAASDGSREVRVGNPLTGGPSMDLI DSTLIQLESSEKSTRK YKSSANACFFVEGHVGSQIGETIENSHPDVTMPANLWASHQAKGTARIFQVSTR EYIRLGLTFEAFGE"

gene	10277..11773
CDS	/gene="phnI" 10277..11773 /gene="phnI" /codon_start=1 /transl_table=11 /product="2-carboxybenzaldehyde dehydrogenase" /db_xref="GI:4586280" /translation="MGSAKSVKKLSGRLMMSSRCRPTCGAIRPAAQHAHFRFOPA RFTVAWAFSQKRLIGSFPMTTEHAYLPFCNLFYGGRFHQGYQTGLNPATQEDLIG PVSNADAVDAVQAQHILAFOWRHTKPAVARAALREFGNLRTHAEELGFDAINC GNPREMAADVHPAATQVDYFAGLVHEAKGTTPMGDGVNLNVSRPYPVVASIVAYN HPLMFALAAGLAPAAVATNTVILKAATAPLSAARLVATLLQDVFPGVNVLSDTPCG DALVDHLPLVKVSLIGSTANGKSIILRSAAEKMPVSLELGKNPLILCEADMNDKALE GAUVGMNFWMAGOSCTSRCFVHRSVDEYVEIKEMQKISQLRYCGLPTDPATMGCL VSKAQYEKMSPFSDSHKECARLVCGVRPSDKALDNMGFVTATVFADVTFAURLFRE EVPFGVLAVIPWDSEESAHHQWAGFRDLDRFNNDKGSRPRSPGISR"
gene	13991..14899
CDS	/gene="phnC" 13991..14899 /gene="phnC" /codon_start=1 /transl_table=11 /product="3,4-dihydroxyphenanthrene dioxygenase" /protein_id="BAA76330.1" /db_xref="GI:4586281" /translation="MANVELTGLVGISVSDAAAANKNYATEVMQMQLVDEGEADRFLYL MDLWHRI VVHTGGDDDLAVGWVRVAGAOLEQMAAKLTAAAGIEVRRGTEABAAERRV LGILLKLQDPAGNAHEILFYSLPVDGHKPEFHGRPMFGGFKTDGOGVGHILIRSYDPRAT MRFEAVLGVTSDDYLLKLTDTGPVRI PELKCNRRHSVALMFGMPKPKRANHLMLVBYQ DMRDLCQTLIDKVHKHGIDITAVQLGVHNSNDGALS FYAAPNSGWLWEFGWSVCAPAQQE
Query Match	10.6%; Score 104.6; DB 1; Length 24448;
Best Local Similarity	49.5%; Pred. No. 4.9e-17;
Matches 328; Conservative	0; Mismatches 329; Indels 6; Gaps 2;
QY	40 TGC GGCTCAGGGAAGCCGTGTTGGCTCAGACCTTGCTAATGGTAGTCGATTCGGTAC 99
Db	2140 TGTTCACTCATGAACGCTACTCGCTGCCGCGCTCAAGCGTGGCTTGGAATTCCTCAT 2199
QY	100 GAGTGTGCATCGGGAGGTTCCGGAGTAGTCAAAATTCAAGTTACTCGAAGGGAATGTC 159
Db	2200 GAATGTCAATCCGGATCGTCAGCAGTTGCCGCTTTTCAGTTGCTGGAGGCGGATGCA 2259
QY	160 TCAATGTGCCGAGTCTCCAGGACITTTCTCGGAGATCGTGAGAAGGCAACCGCCAT 219
Db	2260 GATTGTGTGTAATGCCCGCGGTTTGAATGCCGAGCTCGTGAATCGCGCA--TGCAC 2316
QY	220 CTTGTCATGCCAGTGCCTTGCGCTCTCAGACCTCGGATCAAAGTCGCAGTGCAGGACA 279
Db	2317 CTCGGCTGCCAGAGCACACCAGGTAGTGACTGTGCAATCAAGCTTGAAGCCGGC 2376
QY	280 TAGCTCCCAACGATTCCAATCTCAAGNATGGAGCGGAAGTTGTGAGGTCCGGCGCTA 339
Db	2377 TATGTCCCAGTGGATTCGACGCGCFACAGATGCGGCGTTGGTCGATGTCGTGCCACTG 2436
QY	340 ACTCATGACCTGTGTCGTCCGTCGATACGCACCTGATGGCGCAGCAAAATTCCTCCCG 399
Db	2437 ACGCAAGACATGTTTCGAGTTCCAGTTTCGTACCGAGGCTCCAGCTGAGTTTTTGCCT 2496
QY	400 CAGTTCTGCTAGTAGGGCAGACGATTCGCAAGCGGTGTTTCGCGCATATTCATATGG 459
Db	2497 CAGTATGCGCTGCT---TCGCTTGCCTGGCGTGGTCGGCAACAGGGCGCTACTCGATG 2553
QY	460 AATTTAAAGAACCCCGAAGGCATATGGAGTTCTATATTAAAGAGGTTACCCAAGGACGA 519
Db	2554 AACCTGCCAATGGCATTTGGGAATGGCATTTTCATCGTCAACGCAACCGGGTGGTTGT 2613
QY	520 TTTAGTCTTGGCTTTTCGAAAAATAGAAAAAGAGCGCGCTCGCTATATTTTTGACGGA 579
Db	2614 GGAACCGCGTGTGTTTCGATGTGCTGAACGAGGTGATGCTATATGATTTGAGGTGCT 2673

AUTHORS
TITLE Direct Submission
JOURNAL Submitted (15-NOV-1999) Genetica, Universidad de Sevilla, Ap. 1095, Sevilla 41080, Spain
REMARK
REFERENCE Sequence update by submitter
AUTHORS 8 (bases 1 to 8154)
TITLE Direct Submission
JOURNAL Submitted (01-APR-2002) Departamento de Ciencias Ambientales, Universidad Pablo de Olavide, ctra. Utrera Km 1, Sevilla 41013, Spain
REMARK
REFERENCE Sequence update by submitter
AUTHORS 9 (bases 1 to 8154)
TITLE Direct Submission
JOURNAL Submitted (05-JUN-2002) Departamento de Ciencias Ambientales, Universidad Pablo de Olavide, ctra. Utrera Km 1, Sevilla 41013, Spain
REMARK
COMMENT Sequence update by submitter
ON or before Sep 26, 2002 this sequence version replaced
FEATURES
source
1. 8154
/organism="Spingopyxis macrogoltabida"
/mol_type="genomic DNA"
/strain="TFA"
/db_xref="taxon:33050"
complement (18. .563)
/gene="thnA2"
complement (18. .563)
/gene="thnA2"
/codon_start=1
/transl_table=1
/product="ring hydroxylating dioxygenase beta subunit"
/protein_id="AA26444.1"
/db_xref="GI:23330205"
/db_xref="GI:23330203"
complement (588. .1982)
/gene="thnA1"
complement (588. .1982)
/gene="thnA1"
/codon_start=1
/transl_table=1
/product="ring hydroxylating dioxygenase alpha subunit"
/protein_id="AA26443.1"
/db_xref="GI:23330204"
/translations="METRMVGTGKIEHGAHIEYDSLMNLDGWSVRKVFYDRDIYEE
EMQIFARVWLFVAHESQIPNPGDFVTYMGEDGVIVARQDKSVGFINSCEPHGK
VCADGNARRFYCNHGWAFDIGNLMGNAEEVYDPGDIKSEWQMPVAKVGVYK
GLIFATPDGAPSLDYGDRWYLDICLDSDEGSELVGGTGMKPMI PMNWKVAEEN
VGDAYTAAMVHDSASAKMTGMAFPPIIDFENTFHSSANGHGLEFLGDIGDIALLRP
KVTEYFMDVLKPMKRLGALRSQIFGSIASASVFNLSYLPGISFTFTWLPKPTMT
EVSMTVNFMDPDDIKDEITKAVMQTFGPGCTFENDSDENHEHSTITNKGVYTRQGR
LHVCGGISRKDDAEIPLGIVHRQFADSNQREFHRRWHLDTAESWDDVPDRRSNRLK
GRTRDLPQGLKSL"
complement (1991. .2770)
/gene="thnF"
complement (1991. .2770)
/gene="thnF"
/notes="class II aldolase"
/codon_start=1
/transl_table=1
/product="2,4-dihydroxydec-2-ene-1,10-dioic acid aldolase"
/protein_id="AA26442.1"
/db_xref="GI:23330203"
/translations="MENPONLQALLQKROIGLWLTIGDATSAEICAAGFDMILLI
DGHVHLSLSVLAQIRATSGYPTQAVRVSADPVIIKQVLDMGVQLLVPVQSG
EBAAIYRACRYPPGSGRGVGGARAGWLPDYLKADEQIAIIVQVETRALDDSD
ALMTEGVGVGFI GPADLSASLGLGQPAHPVPAEILHVIDRITRAGKAPGILTRDP
EMLRBATERGALLNAVGLDASLLAAQAKTWKRLMPSAERLT"

gene
CDS
complement (2776. .3579)
/gene="thnE"
complement (2776. .3579)
/gene="thnE"
/codon_start=1
/transl_table=1
/product="2-oxodec-4-ene-1,10-dioic acid hydratase"
/protein_id="AAG18491.2"
/db_xref="GI:23330202"
/translations="MLDDATIRELAAELHAERSGVOTKOFSPDIAFEDSVRVOR
EWEIKLAERKRGKIGKITSRAMOXIANITPDYGVLLDDVIAADSGSVAPGRFTQ
FVUEVAFVINDKLGSPDVSIFDVIGATWVFAVEIIDARIEDFTGSGQRVLD
TIADNAAAGIVGSSPRPDALDRWVGLLKNVVEETGVGAGVLPKAGIIVL
ARKLSAMDEYLAAGELVLAGSFTRIVEVESGQICADYGLGTINRAFE"
complement (3584. .4438)
/gene="thnD"
complement (3584. .4438)
/gene="thnD"
/codon_start=1
/transl_table=1
/product="4-(2-oxocyclohexyl)-2-hydroxy-but-2,4-dienoic
acid hydrolase"
/protein_id="AAG18490.1"
/db_xref="GI:10567587"
/translations="MSEAFVEKLPKPYETHVMIGDPANPLLVHGARPGGSAGS
NRECPADLYKDYVIAHDMTGGQSEIPDPVAHAHTVMGLRIEQAQGVLDGLIEK
THIVGNSGFLALHMLNEQDFDKVTWNGPAGAKMDATPELVNFVADPRLPRY
REVINSFYDPSLVNLEETIQERFKIATDPEVRRVQEVMPESKTIQMPALVPPSAL
KRIPHEVALVEGRHDRVIPLOASLYLLEHLQRAELFVLDRCGHWAQIQRDWMLPIIR
NHFLERAR"
complement (4496. .5308)
/gene="thnB"
complement (4496. .5308)
/gene="thnB"
/codon_start=1
/transl_table=1
/product="cis-dihydrodiol dehydrogenase"
/protein_id="AAN26445.1"
/db_xref="GI:23330206"
/translations="MGWLDGKVALITGGAGIGRAIVERFAEGAKGVLEVNAAGRAE
ELAAVLGDSVCVVTGVSASPADNAAVQRTVEKGLDTPVANAGVDFPQPLAQVSG
DOIGAFDELMSVNVKGLLGRALPVELRSGSVFTTISNAGFYAGGGLPYTASK
HALVGLVQLAHLAPKRVNAVAPGMKTELGLSATGSGSLSQMDGLDTLLKDI
TPLQVSPYQYCGPYVLLASDSDSVTTGAIINTDGGFGVGTGVRGEGEL"
5702. .6616
/gene="thnC"
5702. .6616
/gene="thnC"
/codon_start=1
/transl_table=1
/product="1,2-dihydroxynaphthalene dioxygenase ThnC"
/protein_id="AAD52962.1"
/db_xref="GI:5817852"
/translations="MDKASVTELGVIIGISVSDAAARQDYACCI VGMELVDDGEDRFY
LRVGMHRIIVHTNGGDDLEVLGWRVDPGNELDAIAQLDAAGYSYRVAEAEAAER
RVGLLKLDPGCPNTEIFWGPQETLRPFHPGRPMHGRFITSQGLCHCLLRQDDDE
AAYRFYTLFGGVEYHLFPDQATKPTFMHCNDRQHAFAVAGLGPMPKRNHLMIE
YTELDDIGIAHDMIREROIPVAMOLGKHSNQALTFYHASPFGMWLWELGWGVKSSDQ
LVYYKADLFHGHEASGFLDVELGDDG"
6722. .7036
/gene="thnA3"
6722. .7036
/gene="thnA3"
/codon_start=1
/transl_table=1
/product="Rieske-type ferredoxin"
/protein_id="AAD52963.2"
/db_xref="GI:23330201"
/translations="WGRKVRLCASAEVPPGEMLSVDTDAGDMLAVNVVDGFFVTSNL

Query Match 9.3%; Score 91.6; DB 1; Length 8154;
Best Local Similarity 49.7%; Pred. No. 1.5e-13;
Matches 316; Conservative 0; Mismatches 314; Indels 6; Gaps 3;


```
YLITNIEAFHTDKODEFVVRNFFVYVYHRGQLEHSEHVGCSQDLIRKVGNGFQLARRK
VSLDARVETEDKNLYTF"
5080..5970
/gene="phnC"
CDS
5080..5970
/gene="phnC"
/codon_start=1
/transl_table=11
/product="extradiol dioxygenase"
/protein_id="BAC81543.1"
/db_xref="GI:33456992"
/translation="MVKVIELGYMGFNVTDMAWREYATKCVGLEIREGDRDRFYLR
MDGQHRRTVHOSDEDDLAYAGRWKGPPEEFKAMOOHLTANGVDFRVASBEECEERFV
IGLLKHPSPGNPTLFPSPQMDVMKPPHFGGRPMYKGLTGNQIGIHFIIRSLDDAT
YAFISMLGEGSVETIKLPNGMIAPVFMCHNDQRHSLAFGLGFMKRIINHLMEYT
ELDDGIAHDTARREGIDVAMOLGKHSNDQATFYFFATPSPGWLLELHGAREAVREQE
HYTGDFGHPKPEATLWNEC"
6676..7854
/gene="trpB"
CDS
6676..7854
/gene="trpB"
/codon_start=1
/transl_table=11
/product="tryptophan synthase b subunit"
/protein_id="BAC81544.1"
/db_xref="GI:33456993"
/translation="WPDEKGYFEGYGGVIPPDLVAIMNDINRAVOEISKTPSPKDEL
NDLYHYVGRPSPIFYAKLDEKLAGAKIFLKBEDLNHTGAHKINHCLEALLARHMG
KRYNAETGAGGHALATACALVIGDCEIHEMGVNDIEKERNVTMKILGCTLVPLVS
RQRTILKQVDSAFSEYKDPVNFYIAGSVVGGPPFPMVRDFOSIIGEARAQFLE
EQRULPDTVTCVGGSSNAIGLFTAFDLDELVDIVGVFPAGRGLEGTGHAATMGKVK
GALHPFECYNLODQSGEPLPVYSIASGLDYPGVGPQHLKDLGRKVAIDDOECID
AFMTLSRVEGIIIPALESAAHAYAMKIAKONSNETILVNLSGRDKADFIADKLGL
"
complement(8001..8687)
/gene="orf7"
CDS
complement(8001..8687)
/gene="orf7"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC81545.1"
/db_xref="GI:33456994"
/translation="WNISFFNHQTSIFLQOLNNNRNFWNFKRSLYQSVREPSLAF
IEMEPSIKLSNFTVAAKTGGSLMRIYDARFSSDKTPYKTNIGIOFRTAGKQV
HAFGYLIHTSPPTCFVAGIMRPNKALSIREMISDSFNWKNITRNTFKRYTLLS
GSLKTYPRGYAKDHPMINDLRKDPFIAIHLPLTHEBILSPNLTKTIHRFKVASDLMD
YLCNALLEPV"
8789..9811
/gene="phnA4"
CDS
8789..9811
/gene="phnA4"
/codon_start=1
/transl_table=11
/product="ferredoxin reductase"
/protein_id="BAC81546.1"
/db_xref="GI:33456995"
/translation="WTSYNVKISGQLEFACBEGDTILBAALRAGVGHVPCNSGGCG
ACKVEINEVENIWEDAPGLSPRDIKKGRKLSQCIPTEDELEIKVRLNPEAMPLHKP
IKGKAVLFIINKLTDMAEFCFKTEHPHFAGQFALLDFPGITGSYGMCMNLFNEE
GEWRFIIKKMPDGSATTLFEDYEVGAIEVIDGPYGLAYLKEPIPRDIVCVGGSGLS
PEMSIIKAARDPQLANDRNILFYGGRTPSDIPCKPLIEADDDLGRVKNFNVAVDVE
AAEAGWGDVGDFIHELLGKTLGKLPLEHEFYFCGPPMTDALTRMLMTEYKVPFDQI
HYDRFY"
9907..10221
/gene="phnA3"
CDS
9907..10221
/gene="phnA3"
/codon_start=1
/transl_table=11
/product="ferredoxin"
/protein_id="BAC81547.1"
/db_xref="GI:33456996"
```

```
/translation="MSNIVLCKTDEVKDEAPVAVTPEGPALAVYVFEGEYVYTNLNC
THGMAMLTQGYQDGEVECFPHGAFSIKTGTDPSPQCIPIKTIPTVTVEDGNICIPA
AE"
10236..10580
/gene="phnD"
CDS
10236..10580
/gene="phnD"
/codon_start=1
/transl_table=11
/product="isomerase"
/protein_id="BAC81548.1"
/db_xref="GI:33456997"
/translation="WQESIDFVDFDMXSPSYLAHTOLPDLARKYGLTVKYLPMISIPVA
KTAAGNYGSRNEVPKIKVLMQIDIKWAAHYDVPTFPFNLEIEFPNIGLVLYANDKD
QAEAYVNAAYAKI"
ORIGIN
Query Match 8.6%; Score 85; DB 1; Length 10580;
Best Local Similarity 47.0%; Pred. No. 9.3e-12;
Matches 462; Conservative 0; Mismatches 485; Indels 36; Gaps 5;
Qy 38 CTTGCGGCTCAGGGAAGAGCTGTTGGTCTCAGCACCTTCGTAATGATCGGATTCGGTTC 97
Db 8835 CTTGTGAGGAAGAGATACATATTTTCGTGCGACACTACGCGCAGGTGAGGCATGCCIT 8894
Qy 98 ACAGGTGTGCATCGGAGGTTCGGAGTATGCAAAATTCGAGTTACTCGAAGGAATGTCC 157
Db 8895 ACGAATGTAATTCAGGTGTTGCTGTAAAGCTTGAAGTGTAAACGGTGAAGTGG 8954
Qy 158 AATCAATGTGCGCGATGCTCCAGGACTTCTTCGGGAGATCTGTAGAGGGCAACCGCC 217
Db 8955 AGAATATTTGGAAGATGCGCCGGGCTGTACCTCGTGATATTAATAAAGG---TCGCA 9011
Qy 218 ATCTTGATCGCATCGGTGCTCTCAGACCTGCGGATCAAAATTCGAGTTCGAGTGCAGACA 277
Db 9012 AATTGAGCTGTCAATGATTTCCAACTGAGACCTTGAGATTAGGTTGCTTAAACCCG 9071
Qy 278 AGTAGTCCCAACGATTCCAATCTCAAGAAATGGAAGCGGAAGTTGTGAGGTCCGGCGC 337
Db 9072 AAGCGATCGCGTTACATAAGCAATTAAGGCTAAGGCGCTTTTATTGTAATAATAAAC 9131
Qy 338 TAACATGATGCTGTGCTGCGATTACGCACTGATGCGCGCAGCAAAATTTCTCTCCCG 397
Db 9132 TAACAGAAATATGCGAGAGTTTGTCTTTAAACGAGCATCTGTCTCATTTTAAAGCAG 9191
Qy 398 GCCAGTTTCTCCTAGTAGAGCGAGAGAGTTCGCGCGGTGGTTCGCGCATATTCATAGG 457
Db 9192 GGCAATTTGCTTATTAGATTTCCCGGCGATTACAGGC---TCACGCTGGTTATTCAATGT 9248
Qy 458 CGAATTTAAAGAACCCCGGAGGCATATGGGAGTTCTATATTAAGAGGTTACCCACAGAC 517
Db 9249 GTAACTTGCCAAATGAGGAAGGTGAGTGGGCTTTATTATTAAGAAATGCCAGCGTA 9308
Qy 518 GATTAGTCTCTGGCTTTTCGAAAATAGAAAAGAGCGCTCGTCTATTTTTCAGCGGAC 577
Db 9309 GTGCCACACAACTTTATTTTGAGATTATGAGTGGGTGGGAGATTGTAATCGACGGC 9368
Qy 578 CAATGGGCACATCTTTCTTCGTCAGGAGCCGCCGCGGAGAGTCTTTCGATTCGCGCGC 637
Db 9369 CTTATGTTTGGCATATTTGAACACAGAAATTCAGAGATATTCGTTTGTGTGGTGGTG 9428
Qy 638 GTGCGCGGCTCTCGTATGCGCGCGCTATTGCAACGCGCTCGATGCGCA-----AA 688
Db 9429 GTTCAGGCTTGTACCCGAGATGTCGATCATTAAGGCAGCTGCCAGAGATCTCCAGCTAA 9488
Qy 689 CAGACAAGCCGGTAAAGTTGTTCTACGGCTCAAGAACTCCGCGGAGCGTGTTCGGTGA 748
Db 9489 ATGATAGAAATATTTATTTGTTCTACGAGGTGTCGACCAAGTATATTGTTCGCGCTA 9548
Qy 749 TCGATATCGACATCGATGA-----GGACAAGCTTGAGGTCTCCAGG 790
Db 9549 AGCTTATTGAGCAGATGATGATTATTCGTGCCGAGTGAAGATTTTAAATGCCGTATCAG 9608
```

791	QY	CAGTTACGGAAGACACCGATAGCCTTTGGGAAGGCCCACTGGTTTTATTCATCAGGTTG	850
	Db	ACGTTGAAGCAGCGCGAAGCAGCAGATGGAATGGTGATGTTGGGTTTATCCATGAGTTGT	9668
851	QY	TCGACGCAGCGCTGCTTTGAAACCTACCGGAATACGAAATTTATCTGCGCGTCCACGCG	910
9669	Db	TAGGAAAAACATTGGAGAAAAACTTCCAGACCGAAATTTATTTCTGTGGCCCTCCCTC	9728
911	QY	CTATGCTCGACGCTACTGTGCTCGTATGCTGCT --- CGGCAAGGGTGTTCCACGCGATCAAA	967
9729	Db	CTATGACAGATGCTTTAAACACGATGCTGATGACTGAATACAAAGTGCCATTTGATCAAA	9788
968	QY	TTCAATTTGACGCATTTTTCATAA	990
9789	Db	TTCAATTACGATCGTTTTTATTAA	9811

Search completed: June 27, 2004, 07:56:13
Job time : 4017.03 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 03:11:33 ; Search time 427.15 Seconds
(without alignments)
9846.004 Million cell updates/sec

Title: US-10-658-691-3

Perfect score: 990

Sequence: 1 atgtaccactcaaatga.....atttgacgcatctttctaa 990

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseqn_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	990	100.0	990	2 AAV68080	AAV68080 DNA encod
2	117.2	11.8	5337	3 AAA10957	AAA10957 DNA seque
3	107.8	10.9	5743	3 AAC63945	AAC63945 Burkholde
4	64.6	6.5	1180	4 AAC92962	AAC92962 Pseudomon
5	54.6	5.5	12808	3 AAA65347	AAA65347 NDO relat
6	53	5.4	9706	3 AAA65342	AAA65342 NDO relat
7	49.8	5.0	4355	3 AAA65344	AAA65344 NDO relat
8	45.6	4.6	14462	3 AAA65346	AAA65346 NDO relat
9	43	4.3	2000	7 ADA71938	ADA71938 Rice Gene
10	40.4	4.1	2000	7 ADA71938	ADA71938 Rice Gene
11	39	3.9	4721	3 AAA11712	AAA11712 Comamonas
12	38.4	3.9	7785	4 ABL17653	ABL17653 Drosophil
13	38.4	3.9	12118	4 ABL17652	ABL17652 Drosophil
14	37.6	3.8	680	4 ABL15161	ABL15161 Drosophil
15	37.6	3.8	2738	4 ABL15160	ABL15160 Drosophil
16	37.6	3.8	6763	4 ABL07112	ABL07112 Drosophil
17	37	3.7	1029	7 ACA25179	ACA25179 Prokaryot
18	36.8	3.7	6379	2 AAQ79569	AAQ79569 Nocardia
19	36.8	3.7	6379	2 AAQ79569	AAQ79569 Nocardia
20	36.4	3.7	2394	7 ACA45674	ACA45674 Prokaryot
21	36.4	3.7	7188	7 ACC83205	ACC83205 Methane o
22	35.8	3.6	5331	3 AAA30292	AAA30292 DNA fragm
23	35.6	3.6	632	4 AAI23327	AAI23327 Probe #13

C 24	35.6	3.6	632	4	ABA68435	ABA68435 Human foe
C 25	35.6	3.6	632	4	AAI48649	AAI48649 Probe #17
C 26	35.6	3.6	632	4	ABA50485	ABA50485 Human bre
C 27	35.6	3.6	632	4	ABA35422	ABA35422 Probe #13
C 28	35.6	3.6	632	4	AAK42577	AAK42577 Human bon
C 29	35.6	3.6	632	4	AAK18808	AAK18808 Human bra
C 30	35.6	3.6	632	4	ABS42191	ABS42191 Human liv
C 31	35.6	3.6	632	5	AAI08968	AAI08968 Probe #89
C 32	35.6	3.6	632	6	ABS16628	ABS16628 Human gen
C 33	35.6	3.6	945	7	ACA25001	ACA25001 Prokaryot
C 34	35.6	3.6	1983	4	AAI14117	AAI14117 Probe #40
C 35	35.6	3.6	1983	4	ABA55843	ABA55843 Human foe
C 36	35.6	3.6	1983	4	AAI35499	AAI35499 Probe #41
C 37	35.6	3.6	1983	4	ABA45354	ABA45354 Human bre
C 38	35.6	3.6	1983	4	ABA25515	ABA25515 Probe #39
C 39	35.6	3.6	1983	4	AAK29543	AAK29543 Human bon
C 40	35.6	3.6	1983	4	AAK04063	AAK04063 Human bra
C 41	35.6	3.6	1983	4	ABS29170	ABS29170 Human liv
C 42	35.6	3.6	1983	5	AAI03970	AAI03970 Probe #39
C 43	35.6	3.6	1983	6	ABS04097	ABS04097 Human gen
C 44	34.8	3.5	65042	7	ABS55711	ABS55711 DNA encod
C 45	34.8	3.5	65042	8	ACA63029	ACA63029 Human gen

ALIGNMENTS

RESULT 1
AAV68080
ID AAV68080 standard; DNA; 990 BP.

XX
AC AAV68080;
DT 12-JAN-1999 (first entry)
XX
DE DNA encoding an oxidase.
XX
KW Oxidase; dioxin removal; ds.
OS Pseudomonas sp.
XX
FH Key Location/Qualifiers
FT CDS 1..990
FT /tag= a
FT /product= "oxidase"
XX
PN JP10257895-A.
XX
PD 29-SEP-1998.

XX
PF 18-MAR-1997; 97JP-00084401.
XX
PR 18-MAR-1997; 97JP-00084401.
XX
PA (ASAH) ASAH KASEI KOGYO KK.
XX
DR WPI; 1998-575906/49.
XX
DR P-PSDB; AAW80338.
XX
PT Oxidase gene derived from a microbe for removal of dioxin - converts hetero:poly:cyclic aromatic hydrocarbon to a benzoic acid homolog.
XX
PS Claim 2; Page 12-13; 15pp; Japanese.

XX
CC The present sequence encodes an oxidase which is derived from a microbe and converts a heteropolycyclic aromatic hydrocarbon to a benzoic acid homologue. The oxidase is used in a method for the removal of dioxin in which dibenzo-p-dioxin is converted to a diphenyl ether triol

XX
SQ Sequence 990 BP; 230 A; 248 C; 279 G; 233 T; 0 U; 0 Other;
Query Match 100.0%; Score 990; DB 2; Length 990;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 990; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGTACCAACTCAAAATTAAGGCGAAGCGCAGAGACTCGGCTCAGGAGAGCGCTG 60
Db 1 ATGTACCAACTCAAAATTAAGGCGAAGCGCAGAGACTCGGCTCAGGAGAGCGCTG 60

Qy 61 TTGGTCTCAGCACTTGCTAATGGTATCGGATTCGGTACGAGTGTGCATCGGAGGTTC 120
Db 61 TTGGTCTCAGCACTTGCTAATGGTATCGGATTCGGTACGAGTGTGCATCGGAGGTTC 120

Qy 121 GGAGTATGCAAAATTCGAGTTACTCGAAGGGATGTCATCAATCAATGCGCGAGTCTCA 180
Db 121 GGAGTATGCAAAATTCGAGTTACTCGAAGGGATGTCATCAATCAATGCGCGAGTCTCA 180

Qy 181 GGACTTTCTTCGAGATCGTGAGAGGCAACCGCCATCTTGCATGCGAGTGTGCG 240
Db 181 GGACTTTCTTCGAGATCGTGAGAGGCAACCGCCATCTTGCATGCGAGTGTGCG 240

Qy 241 CTTCTAGACCTCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCG 300
Db 241 CTTCTAGACCTCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCGGATCG 300

Qy 301 TCAAGAATGGAAGCGGAGTGTGAGGTCGCGGCTAACTCATGACTGTGTCGCTG 360
Db 301 TCAAGAATGGAAGCGGAGTGTGAGGTCGCGGCTAACTCATGACTGTGTCGCTG 360

Qy 361 CGATTACGACCTGATGCGGACGAAATTTCTTCCCGGCGAGTTCTGCTAGTAGAGCA 420
Db 361 CGATTACGACCTGATGCGGACGAAATTTCTTCCCGGCGAGTTCTGCTAGTAGAGCA 420

Qy 421 GAGCAGTTCCAGGCGTGTGTCGCAATTCATGCGGCTAACTCATGACTGTGTCGCTG 480
Db 421 GAGCAGTTCCAGGCGTGTGTCGCAATTCATGCGGCTAACTCATGACTGTGTCGCTG 480

Qy 481 ATATGGAGTCTTATTAAGAGGATGACCAAGAGGATTTAGTCTGCTGCTTTTCA 540
Db 481 ATATGGAGTCTTATTAAGAGGATGACCAAGAGGATTTAGTCTGCTGCTTTTCA 540

Qy 541 AATAGAAAAGAGGCGCTGCTTATTTTGGAGGACCAATGCGGACATCTTCTTCGCT 600
Db 541 AATAGAAAAGAGGCGCTGCTTATTTTGGAGGACCAATGCGGACATCTTCTTCGCT 600

Qy 601 CCAGGACCGCGCGAAGAGTCTTTCATGTCGCGGCTGCGGCTCTCGTATGCGGCC 660
Db 601 CCAGGACCGCGCGAAGAGTCTTTCATGTCGCGGCTGCGGCTCTCGTATGCGGCC 660

Qy 661 GCTATTGCGCGCTCGATGCGGACGACAGAGCGGTAAGTCTTCTACGGCTCA 720
Db 661 GCTATTGCGCGCTCGATGCGGACGACAGAGCGGTAAGTCTTCTACGGCTCA 720

Qy 721 AGAATCCCGCGAGCTGTTTCGGTGGATCGATATCGACATCGATGAGGACAAAGCTTGA 780
Db 721 AGAATCCCGCGAGCTGTTTCGGTGGATCGATATCGACATCGATGAGGACAAAGCTTGA 780

Qy 781 GTCTCCAGGAGTACGAGACGAGTACGCTTTCGAGGCGGCTGCTGCTGCTGCTGCTGCT 840
Db 781 GTCTCCAGGAGTACGAGACGAGTACGCTTTCGAGGCGGCTGCTGCTGCTGCTGCTGCT 840

Qy 841 CATCAGGTTGTGAGCGAGCGCTGCTTGAACCCCTACCGGATACGAAATTTATCTTCC 900
Db 841 CATCAGGTTGTGAGCGAGCGCTGCTTGAACCCCTACCGGATACGAAATTTATCTTCC 900

Qy 901 GGTCCAGCGCTATGTCAGCGTACTGTCGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db 901 GGTCCAGCGCTATGTCAGCGTACTGTCGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 960

Qy 961 GATCAAAATTCATTTTACGCAATTTTCTAA 990
Db 961 GATCAAAATTCATTTTACGCAATTTTCTAA 990

RESULT 2
AAA10957

AAA10957 standard; DNA; 5337 BP.
AAA10957;
14-JUL-2000 (first entry)
DNA sequence of enzyme with phenylacetic acid hydroxylase activity.
Phenylacetic acid hydroxylase; prepare; m-hydroxyphenylacetic acid; drug;
agricultural chemical; ds.
Pseudomonas aeruginosa.
JP2000069974-A.
07-MAR-2000.
01-SEP-1998; 98JP-00247139.
01-SEP-1998; 98JP-00247139.
(SUMO) SUMITOMO CHEM CO LTD.
WPI; 2000-264450/23.
Preparation of m-hydroxyphenylacetic acid - used as a raw material for
drugs and agricultural chemicals.
Claim 3; Page 4-6; 6pp; Japanese.
This sequence represents the nucleotide sequence of an enzyme with
phenylacetic acid hydroxylase activity, which is used in the method of
the invention. The invention relates to a method for the preparation of m
-hydroxyphenylacetic acid. The method comprises the use of a culture
liquid, microbe body or a treated product of a microbe body, derived from
pseudomonas aeruginosa. The microbe produces an enzyme with the ability
of hydroxylating the meta site of phenylacetic acid, which gives rise to
the production of m-hydroxyacetic acid. The m-hydroxyacetic acid is used
as a raw material for drugs and agricultural chemicals
Qy Sequence 5337 BP; 1135 A; 1512 C; 1672 G; 1018 T; 0 U; 0 Other;
Query Match 11.8%; Score 117.2; DB 3; Length 5337;
Best Local Similarity 52.8%; Pred. No. 1.6e-27;
Matches 325; Conservative 0; Mismatches 283; Indels 8; Gaps 3;
Qy 58 CTGTTGCTCTCAGCACTTGTCTAATGCTATCGGATTTCCGTACGAGTGTGCATCGGAGGT 117
Db 4283 CTCTTGGGTCGGCGCTCCAGGCGGCTGTCGGGTTCCTGTCAGAGTSCCATTCGGGGGGC 4342
Qy 118 TCGGAGTATGCAAAATTCGAGTTACTCGAAGGGAATGTCCAATCAATGTGGCCGATGCT 177
Db 4343 TCGGAGTATGCGCTTCGAGTTGTTGAGGGGGAATGCGATGCTGTTGTG--CCGAGCC 4400
Qy 178 CAGGACTTCTTCGGGAGATCGTGAGAGGGAACCGGCACTTCGATCGCAGTGGT 237
Db 4401 CCGGCGCTTACCGACCGGTGACCGACGCC--GCAACCGCTATCTAGCTGTCTAGTGTGTC 4457
Qy 238 GCGCTCTCAGACTCGGATCAAGTTCGAGTGCAGTGCAGGACAAGTACGTCCCAACGATTCCA 297
Db 4458 GCCAAGACCGACCTTCGGATCAAGTGAAGCGCGCTGAGGAATATCGCCCCAGGTACAG 4517
Qy 298 ATCTCAAGATGGAAGCGGAAGTTGTGAGGTCCGGGCGCTAACTCATGACCTGTGTCC 357
Db 4518 CCGCGGTATTTGCTGCGCGGTTCGAGGCGCTTACGGGTAAACGATGACATCTTGGAG 4577
Qy 358 GTGCGATTACGACTGATGGCCAGCAAAATTTCTCCCGCCGAGTTCGCTTAGTAGAG 417
Db 4578 TTGCGTTCTCGTGTATGSCCTTCGCGAGTTCCTCCCGGCGAGTACGGCTGAT--G 4634
Qy 418 GCAGAGCAGTTGCCAGCGTGGTTCGCGCATATTTCAATGCGGAATTTAAAGAACCCCGAA 477
Db 4635 CAATGCTGCTGGTGGAGGTTTCGCGGCTTACTCGATGTCTCAACACGCGCAACCTCGG 4694

QY	478	GGCATATGGGAGTTCTATATTAAGAGGGTACCCACAGACGATTTAGTCTCTTGGCTTTTC	537		
Db	4695	GGAGAGTGGCATTTTCCAGATCCGGCGGGTGCCCAATGCGAAAGACGACTACCTGGCC	4754		
QY	538	GAATAATAGAAAGAGCGCTCGTCTATTTTGAACGGGACCAATGGGCACATCTTTCTTC	597		
Db	4755	GGGCATCTCAAAACAGGGATTCGTCGAGGTTGAGGGCCCTATGSCCTTGCTATTG	4814		
QY	598	CGTCCAGGGACCGGCCGAAGAGTCTTTGCTATGGCGCGGTGCGGGCTCTCGTATGCG	657		
Db	4815	AGAACCGAAAGTCCGGCGGACATCGTCTGCTGCGGAGCTCGGGGCTCGCGCCTGTG	4874		
QY	658	GCCGCTATTGCACGCG	673		
Db	4875	GTGTCCATCGCGCGTG	4890		
RESULT 3					
AA63945	AAC63945 standard; DNA; 5743 BP.				
XX	AAC63945;				
XX	15-SEP-2003 (revised)				
DT	09-FEB-2001 (first entry)				
XX	Burkholderia sp. C1 desulphurase gene cluster.				
DE	Desulphurase gene cluster; DBT dioxygenase; ferredoxin subunit;				
KW	Fe-S alpha subunit; dihydrodiol DBT dehydrogenase;				
KW	cis 3-hydroxythianaphenyl 2-oxo-3-butenate isomerase;				
KW	fossil fuel cleanup; ds.				
XX	Burkholderia sp; C1.				
OS	Burkholderia sp; C1.				
XX	Burkholderia sp; C1.				
FH	Key	Location/Qualifiers			
FT	CDS	78..1106			
FT	FT	/*tag= a			
FT	FT	/product= "DBT dioxygenase ferredoxin subunit (AAB29472)"			
FT	FT	2499..3332			
FT	FT	/*tag= b			
FT	FT	/product= "Dihydrodiol DBT dehydrogenase (AAB29473)"			
FT	FT	3389..4747			
FT	FT	/*tag= c			
FT	FT	/product= "DBT dioxygenase Fe-S alpha subunit (AAB29474)"			
FT	FT	4777..5274			
FT	FT	/*tag= d			
FT	FT	/product= "DBT dioxygenase Fe-S alpha subunit (AAB29475)"			
FT	FT	5455..5745			
FT	FT	/*tag= e			
FT	FT	/product= "cis 3-hydroxy-thianaphenyl 2-oxo-3-butenate isomerase (AAB29476)"			
XX	JP2000245477-A.				
PN	XX				
XX	XX				
PD	12-SEP-2000.				
XX	XX				
PF	04-MAR-1999;	99JP-00057446.			
XX	XX				
PR	04-MAR-1999;	99JP-00057446.			
XX	XX				
PA	(SEKI-) ZH SEKIYU SANGYO KASSEIKA CENTER.				
XX	XX				
XX	WPI; 2000-622120/60.				
DR	P-PSDB; AAB29472, AAB29473, AAB29474, AAB29475, AAB29476.				
XX	XX				
PT	Gene encoding ferredoxin reductase subunit of DBT dioxygenase for				
FT	removing sulfur components contained in fossil fuel.				
XX	XX				
PS	Claim 1; Page 9-15; 24pp; Japanese.				
XX	XX				
CC	The invention relates to proteins from Burkholderia sp. C1 which have				

CC	desulphurase activity (AAB29472-B29476), and to the genes encoding them
CC	(desulphurase gene cluster shown in AAG63945). The proteins of the
CC	invention are a ferredoxin reductase subunit of DBT dioxxygenase
CC	(AAB29472), a dihydroadiol DBT dehydrogenase (AAB29473), Fe-S alpha
CC	subunits of DBT dioxxygenase (AAB29474, AAB29475) and a cis 3-hydroxy-
CC	thianaphenyl 2-oxo-3- butenolate isomerase (AAB29476). The invention also
CC	relates to variants of the desulphurase proteins which retain activity,
CC	and to constructs and transformants comprising DNA encoding the
CC	desulphurase proteins. The desulphurase proteins can be used for removing
CC	sulphur compounds from fossil fuels. The present sequence represents the
CC	Burkholderia sp. C1 desulphurase gene cluster. (Updated on 15-SEP-2003 to
CC	standardise OS field)
XX	
SQ	Sequence 5743 BP; 1240 A; 1473 C; 1648 G; 1382 T; 0 U; 0 Other;
	Query Match 10.9%; Score 107.8; DB 3; Length 5743;
	Best Local Similarity 49.8%; Pred. No. 2.2e-24;
	Matches 330; Conservative 0; Mismatches 327; Indels 6; Gaps 2;
QY	40 TCGGGCTCAGGGAAGACCTGTGGTCTCAGCACTTGCTAATGTTATCGGATTCCTCGTAC 99
DB	135 TGTTCCTACTGATGAAGACGCTACTCGCTGCGCGCTCAAGGCTGGCGTTGGATTTCCCTAT 194
QY	100 GAGTGTGCATCGGGAGGTTCGGGAGTATGCAAAATTCGAGTTTACTCGAAGGGAATGTCCAA 159
DB	195 GAATGTCAATCCGATCTGTCAGCAGATTCGCGCTTCAGTTGCTGGAGGGCGATGTCAAG 254
QY	160 TCAATGGCGCGATGCTCCAGGACTTTCTTCGGCAGATCGTAGAAGGGCAACCGCCAT 219
DB	255 GATTGTGGTTCGAATGGCGCGGTTTGAATGCCGAAGCTCGTAATGCGGCA---TGCAC 311
QY	220 CTTCGATCCAGTSCGTTGGCTCTCAGACCTCGCGGATCAAAGTCGCAAGTCGAGACAAG 279
DB	312 CTCGGCTCCAGACACACACGAGTAGTGACTGTGCAATCAAGCTGCGCTTGAAGCGCGGC 371
QY	280 TACGTCCCAAGCATCCAAATCTCAAGAAATGGAACCGGAAGTTGTGTAGGTCCGGCGGCTA 339
DB	372 TATGTCGCGCAGTGTGATCCGACCGCTCAGCATCGCGGTTGGTTCGATGTCGTGCCACTG 431
QY	340 ACTCATGACCTGCTGTCGGTCGATTAACGCACTGATGGCCAGCAAAATTTCTCTCCCGCGC 399
DB	432 ACGCAAGCATGTTTCGAGTTCAGTTCGATACGAGGCTCCAGCTGAGTTTTCCTGCTGCG 491
QY	400 CAGTTCTGCTAGTAGAGGCAGACAGTTCGCCAGCGGTGTTCCGCNCATATTCATGCGG 459
DB	492 CAGTATGCGTGTCT---TCGCTTCGCTGGGTGTCGCCACAAGGGCGGTACTCTCATGAGC 548
QY	460 AATTAAAGAACCCCGAAGGCATATGGAGTTCCTATATTAAGAGGGTACCCACAGACGA 519
DB	549 AACCTGCCAATGSCATTTGGCGAATGSCATTTTCATCGTCAAAACGAAACCGGGTGGTTGT 608
QY	520 TTTAGTCTTGGCTTTTCGAAAATAGAAAAGAGCGCTCGTCTATTTTTCACGGGACCA 579
DB	609 GGAACCGCGGTCTTTTCGATGTCTCCTGAAGCGAGGTGATGCTATTGAGTTGGAAGGGTCT 668
QY	580 ATGGGCACATCTTTCTTCGCTCCAGGACACGGCCGCGAAAAGATCTTTTGCAATGGCGCGGT 639
DB	669 TACGGACCGCTTATCTCGGTACCGAAACCGTAGAGGTGTGCTCTGCATTTGGTGGCGGC 728
QY	640 GCCGGGTCTGTATGGCGCGCTTATTGACGCGCCTCGATTCGCGGAAAACAGACAGCGG 699
DB	729 TCGGAGTTTCTCCGATTTGTCTCAATCCTGCGCGCGCAGTGTGCAATCCGCGCAATGACC 788
QY	700 GTA 702
DB	789 GAA 791
RESULT 4	
AAC92962	
ID	AAC92962 standard; DNA; 1180 BP.
XX	
AC	AAC92962:

RESULT 4
AAC92862
ID AAC92962 standard; DNA; 1180 BP.
XX
AC AAC92962:

XX 27-MAR-2001 (first entry)
XX Pseudomonas mendocina KR1 tmo F gene.
XX Monooxygenase; epoxide preparation; enantio-specific;
XX non-haem diiron-containing monooxygenase; alkene epoxidation; bioreactor;
XX pharmaceutical production; toluene monooxygenase;
XX toluene-4-monooxygenase; T4MO; tmo F; ds.
XX Pseudomonas mendocina.
XX WO200073425-A1.
XX 07-DEC-2000.
XX 26-MAY-2000; 2000WO-US014637.
XX 28-MAY-1999; 99US-0136602P.
XX (ENVI-) ENVIROGEN INC.
XX Steffan RJ, Mcclay KR;
XX WPI; 2001-049935/06.
XX P-PSDB; AAB60228.
XX Preparation of enantio-specific epoxides useful in synthetic organic
XX chemistry and pharmaceutical reactions, by contacting alkene with an
XX enzyme comprising a native or mutated non-haem diiron-containing
XX monooxygenase.
XX Example 3; Fig 16B; 70pp; English.
XX The invention relates to a method for the enantio-specific preparation of
XX epoxides using a wild-type or a mutant non-haem diiron-containing
XX monooxygenase, and an alkene substrate. In particular, the non-haem
XX diiron-containing monooxygenase is a toluene monooxygenase. The invention
XX also relates to mutant non-haem diiron-containing monooxygenases which
XX produce a different ratio of (R) and (S) enantiomers of an epoxide
XX relative to the wild-type enzyme, and methods of producing such mutant
XX enzymes. The method is used for large scale production of epoxides which
XX are useful in many processes, particularly in synthetic organic chemistry
XX and in the production of pharmaceuticals. In an exemplary approach,
XX alkenes are passed through a fluid-bed reactor inoculated with bacteria
XX such as Pseudomonas mendocina KR1 possessing or transformed with a
XX mutated non-haem diiron-containing monooxygenase. The present sequence
XX represents the Pseudomonas mendocina KR1 toluene-4-monooxygenase (T4MO)
XX tmo F subunit gene
XX Sequence 1180 BP; 300 A; 256 C; 311 G; 313 T; 0 U; 0 Other;
XX
XX Query Match 5.5%; Score 64.6; DB 4; Length 1180;
XX Best Local Similarity 51.0%; Pred. No. 2.5e-10;
XX Matches 179; Conservative 0; Mismatches 169; Indels 3; Gaps 1;
XX
XX 68 CAGCAGCTGCTAATGGTATCGGATTTCCGTACAGAGTGTCATCGGAGGTGGGAGTAT 127
XX 119 CTGCTACGTGCTGAATGGTATTTCCATATGAGTGTAATCAGAGGTGGGCGCAT 178
XX 128 GCATATTCGAGTACTCGAAGGATGTCCAATCAATGTGCGCGATGCTCCAGGACTTT 187
XX 179 GTAAGATCGAGTCTTGAGGGGAGAGTCTTAACCTATGCGCTGATGCACAGGATTAG 238
XX 188 CTTCCGCGAGATCGTGAGAAGGGAACCGCATCTTGATCCAGTGGTGTGCGCTCTCAG 247
XX 239 CCGCCCTGACTCCGTAAG--AATCGTTTTTGGCGTCCAGTGCACACATTTATCG 295
XX 248 ACCTGCGGATCAAGTCCGAGTGCAGGACGAAGTACGTCCCAAGATTCCAAATCTCAAGAA 307
XX 296 ACCTCAAAATTAAGGTATTAAACCGTCCGGAGGAGCGTGTTCACATCCCCCAACGTT 355
XX 308 TGGAAGCGGAAGTTGTTGAGTCCGGCGGCTAACTCATGACCTGCTCGTGGATTAC 367

Db 356 TCTCGACTCGAGTAGTTAGTAAGCGCTTCTCTCTGACGAGATGTTGAGTGGACTTG 415
QY 368 GCACCTGATGGCGCCAGCAAAATTTCTCCCGCCGAGTTCTGCTAGTAGAGG 418
Db 416 AAGCGGAACAGAAAGTGCTGTTTTCACACAGGCGCAATATTTATGTTGACG 466
RESULT 5
ID AAA65347 standard; DNA; 12808 BP.
XX AAA65347;
XX AC AAA65347;
XX DT 09-NOV-2000 (first entry)
XX DE NDO related complex alpha subunit nucleotide sequence SEQ ID NO:10.
XX KW Pseudomonas sp. strain NCIB 9816-4; naphthalene dioxygenase; NDO;
XX inducible multi-component enzyme; alpha subunit; mutant; chiral diol;
XX polymer; resin; pharmaceutical; rubber industry; bioremediation; ds.
XX OS Pseudomonas sp.
XX Synthetic.
XX PN WO200037480-A1.
XX PD 29-JUN-2000.
XX PF 26-OCT-1999; 99WO-US025079.
XX PR 26-OCT-1998; 98US-0105575P.
XX PS (IOWA) UNIV IOWA RES FOUND.
XX PI Parales R, Gibson D, Resnick S, Lee K;
XX WPI; 2000-452174/39.
XX P-PSDB; AAB12573.
XX Novel naphthalene dioxygenase mutant having a specific amino acid
XX substitution for preparing chiral diols for use in the polymer, resin,
XX pharmaceutical or rubber industry and for carrying out bioremediation.
XX Disclosure; Page 81-86; 151pp; English.
XX The present invention describes a naphthalene dioxygenase (NDO) or NDO
XX related complex (I) comprising several polypeptides which contain an
XX alpha subunit that contains substituted amino acids at specific
XX positions. The polypeptides and host cells are useful for preparing (-)-
XX (1S,2R)-cis-naphthalene dihydrodiol, (- or +)-cis-biphenyl-3,4-
XX dihydrodiol, (1S,2R)-cis-phenanthrene-1,2-dihydrodiol, cis-1,2-
XX dihydroxyindan, 1,2-dihydroxy-1,2,3,4-tetrahydronaphthalene, which
XX involves contacting them with naphthalene, biphenyl, phenanthrene,
XX indene, and 1,2-dihydronaphthalene respectively. The polypeptides and the
XX host cells are also useful for preparing 1,2-dihydroxy-1,2-
XX dihydrophenanthrene or 3,4-dihydroxy-3,4-dihydrophenanthrene which
XX involves contacting them with phenanthrene. The polypeptides and the host
XX cells are also used in bioremediation in which they oxidize an aromatic
XX compound such as indene, 1,2-dihydronaphthalene, benzocyclopent-1-ene,
XX anthracene, phenanthrene, dibenzo(1,4)dioxan, acenaphthylene,
XX naphthalene, biphenyl, fluorene, dibenzofuran, benzothiofene, 9,10-
XX dihydroanthracene, or 9,10-dihydrophenanthrene a corresponding dihydro
XX dihydroxy compound. The polypeptides and the host cells are useful for
XX preparing chiral diols for use in the polymer, resin, pharmaceutical or
XX rubber industry. The present sequence encodes a naphthalene dioxygenase
XX (NDO) mutant alpha subunit, given in the exemplification of the present
XX invention
XX Sequence 12808 BP; 2992 A; 3328 C; 3576 G; 2912 T; 0 U; 0 Other;
XX

Query Match 5.5%; Score 54.6; DB 3; Length 12808;
Best Local Similarity 50.4%; Pred. No. 2e-06;

Matches	191;	Conservative	0;	Mismatches	179;	Indels	9;	Gaps	2;
Qy	303	AAGAATGGAAGCGGAAGTTGTGAGGTCGGGGCGCTAACTCATGACCTGCTGTCGGCGG	362						
Db	1101	AATCATCAAAAGGCATGTTGTGCGGTCGAGTCGCCACACACATATCGCCGATCCG	1160						
Qy	363	ATTACGCACATGATGGCCAGCAAAATTTCTCTCCCGCGCAGTTCTGCCTAGTAGAGGAGA	422						
Db	1161	CATTGCGCTCGCAAGCCCTTCGAGTTCTCACCGGCAGTAC-----GCGATGCTACA	1214						
Qy	423	GCATTTCCCAAGCGCTGTTGCGGCATATTCAATGGCGCAATTTAAAGAACCCCGAAGGCAT	482						
Db	1215	GTTCAAGTCCCGAACATGTGCGTCATATTCAATGGCTGGTCTGCCAGATGACCAAGAAAT	1274						
Qy	483	ATGGAGATTCTATATTAAAGAGGTACCCACAGSACGATTTAGTCTTTGGCTTTTCGAAA	542						
Db	1275	--GGAGTTCCATATCCGCAAGTCCGGGGCGCGTGTACCGAGTATATTTTCGAGCA	1331						
Qy	543	TAGAAAAGAGGCGCTCGTCTATTTTTCACGGGACCAATGGGCACATCTTCTTCGCTCC	602						
Db	1332	CGTCCGGAAGGTACAAGCATTAAGTTGAGTGGGCGCTCTTGGTACGGCCCTATCTGCGTCA	1391						
Qy	603	AGGGACCGGCGCAAGAGTCTTTGCATTTGGCGGCGGTCCGGGGTCTCGTAGGGGCGCG	662						
Db	1392	GGCTCACACGGGCGGATGCTGTGTGGGCGGCGGACCGGACTCGACCGGTGCTGTC	1451						
Qy	663	TATTGCAGCGGCGCTCGATG	681						
Db	1452	GAITGTTGCGGGCGCGCTG	1470						

RESULT 6
AAA65342
ID AAA65342 standard; DNA; 9706 BP.
XX AC
XX AAA65342;
DT 09-NOV-2000 (first entry)
DE NDO related complex alpha subunit nucleotide sequence SEQ ID NO:5.
XX
XX Pseudomonas sp. strain NCIB 9816-4; naphthalene dioxygenase; NDO;
KW inducible multi-component enzyme; alpha subunit; mutant; chiral diol;
KW polymer; resin; pharmaceutical; rubber industry; bioremediation; ds.
XX
OS Pseudomonas sp.
OS Synthetic.
XX WO200037480-A1.
PN 29-JUN-2000.
XX
XX 26-OCT-1999; 99WO-US025079.
PF
XX 26-OCT-1998; 98US-0105575P.
PR
XX (IOWA) UNIV IOWA RES FOUND.
XX
XX Parales R, Gibson D, Resnick S, Lee K;
PI WPI: 2000-452174/39.
DR P-PSDB; AAB:2568.
DR
XX Novel naphthalene dioxygenase mutant having a specific amino acid
FT substitution for preparing chiral diols for use in the polymer, resin,
PT pharmaceutical or rubber industry and for carrying out bioremediation.
PT
XX Disclosure; Page 66-70; 151pp; English.
XX
XX The present invention describes a naphthalene dioxygenase (NDO) or NDO
CC related complex (I) comprising several polypeptides which contain an
CC alpha subunit that contains substituted amino acids at specific
CC positions. The polypeptides and host cells are useful for preparing (-)-

```

CC (1S,2R)-cis-naphthalene dihydrodiol, (- or +)-cis-biphenyl-1,3,4-
CC dihydrodiol, (1S,2R)-cis-phenanthrene-1,2-dihydrodiol, cis-1,2-
CC dihydroxyindan, 1,2-dihydroxy-1,2,3,4-tetrahydronaphthalene, which
CC involves contacting them with naphthalene, biphenyl, phenanthrene,
CC indene, and 1,2-dihydronaphthalene respectively. The polypeptides and the
CC host cells are also useful for preparing 1,2-dihydroxy-1,2-
CC dihydropentanethrene or 3,4-dihydroxy-3,4-dihydropentanethrene which
CC involves contacting them with phenanthrene. The polypeptides and the host
CC cells are also used in bioremediation in which they oxidize an aromatic
CC compound such as indene, 1,2-dihydronaphthalene, benzocyclohept-1-ene,
CC anthracene, phenanthrene, dibenzo(1,4)dioxan, acenaphthylene,
CC naphthalene, biphenyl, fluorene, dibenzofuran, dibenzothiophene, 9,10-
CC dihydroanthracene, or 9,10-dihydronaphthalene a corresponding dihydro
CC dihydroxy compound. The polypeptides and the host cells are useful for
CC preparing chiral diols for use in the polymer, resin, pharmaceutical or
CC rubber industry. The present sequence encodes a naphthalene dioxygenase
CC (NDO) mutant alpha subunit, given in the exemplification of the present
CC invention
XX
SQ Sequence 9706 BP; 2331 A; 2528 C; 2629 G; 2218 T; 0 U; 0 Other;

Query Match          5.4%; Score 53; DB 3; Length 9706;
Best Local Similarity 50.1%; Pred. No. 6e-06;
Matches 190; Conservative 0; Mismatches 180; Indels 9; Gaps 2;

QY 303 AAGATCGAAGCGGAAGTTCTTGAGTCCGGCGCTAACTCATGACCTGTGTCCGTGCG 362
DB 1193 AATCATCGAGGCGACAGTGTGCGAGTCGAGTCGCCCACTCAGCATATCGTTCGTTCAGC 1252

QY 363 ATTACGCACATGATGGGCGACCAATTTCTTCCCGGCGCAGTTCTGCTAGTAGAGGCAGA 422
DB 1253 CGTAGCGCTCTCAAGGCCCTTCGAGTTCTCACCGGACAGTACGGGACACT-----GCA 1306

QY 423 GCAGTTGCCAGGCGTGTGTCGCGCATATTTCAATGGCGGAATTTAAGAACCCCGHAGGCAT 482
DB 1307 GTTCAGCGCTTGACATCGCGGTCCGTTATTCATATGGCAGGTTTGCAGATGACCAAGAAAT 1366

QY 483 ATGGAGTCTTATATTAAGAGGTTACCCACAGACAGATTTAGTCTTTGGCTTTTCGAAA 542
DB 1367 ---GGAGTTCCACATACGCAAGTTCGCGGTGGCGCGTCACGGAGTATGTTTCGAACA 1423

QY 543 TAGAAAAGAAGCGCTCGTCTATTTTTCGAGGACCAATGGGCACATCTTTTTCCTCGCTCC 602
DB 1424 CGTCCGGAAGGTACAGCATCAAGTTGAGCGGCGCTCTTTGGTACGGCTTATCTACGTCA 1483

QY 603 AGGACCGGCGGAAAGAGTCTTTGCATTTGCGCGGTGCGGGGTCTCGTATGCGGCGCG 662
DB 1484 GAGCACACCGGACCGATGCTGTGTGTAGGTGCGGGACCGGACTCGCACCGGTGCTGTC 1543

QY 663 TATTGCACGCGCTCGATG 681
DB 1544 GATTTGCGGGCGCGCTG 1562

```

RESULT 7	
AAA65344	
ID	AAA65344 standard; DNA; 4355 BP.
XX	
XX	
AC	AAA65344;
XX	
DT	09-NOV-2000 (first entry)
XX	
DE	NDO related complex alpha subunit nucleotide sequence SEQ ID NO:7.
XX	
XX	
KW	Pseudomonas sp. strain NCIB 9816-4; naphthalene dioxygenase; NDO;
KW	inducible multi-component enzyme; alpha subunit; mutant; chiral diol;
KW	polymer; resin; pharmaceutical; rubber industry; bioremediation; ds.
XX	
XX	
OS	Pseudomonas sp.
OS	Synthetic.
XX	
PN	WC2000037480-A1.
XX	
XX	

PD 29-JUN-2000.
PF 26-OCT-1999; 99WO-US025079.
XX 26-OCT-1998; 98US-0105575P.
PR (IOWA) UNIV IOWA RES FOUND.
XX Parales R, Gibson D, Resnick S, Lee K;
PI WPI; 2000-452174/39.
DR P-PSDB; AAB12570.
XX Novel naphthalene dioxygenase mutant having a specific amino acid
PT substitution for preparing chiral diols for use in the polymer, resin,
PT pharmaceutical or rubber industry and for carrying out bioremediation.
XX Disclosure; Page 71-73; 151pp; English.
PS
XX The present invention describes a naphthalene dioxygenase (NDO) or NDO
CC related complex (I) comprising several polypeptides which contain an
CC alpha subunit that contains substituted amino acids at specific
CC positions. The polypeptides and host cells are useful for preparing (-)-
CC (1S,2R)-cis-naphthalene dihydrodiol, (- or +)-cis-biphenyl-3,4-
CC dihydrodiol, (1S,2R)-cis-phenanthrene-1,2-dihydrodiol, cis-1,2-
CC dihydroxyindan, 1,2-dihydroxy-1,2,3,4-tetrahydronaphthalene, which
CC involves contacting them with naphthalene, biphenyl, phenanthrene,
CC indene, and 1,2-dihydronaphthalene respectively. The polypeptides and the
CC host cells are also useful for preparing 1,2-dihydroxy-1,2-
CC dihydrophenanthrene or 3,4-dihydroxy-3,4-dihydrophenanthrene which
CC involves contacting them with phenanthrene, biphenyl, phenanthrene,
CC anthrane, phenanthrene, dibenzo(1,4)dioxan, acenaphthylene, 9,10-
CC dihydroanthracene, or 9,10-dihydrophenanthrene a corresponding dihydro
CC dihydroxy compound. The polypeptides and the host cells are useful for
CC preparing chiral diols for use in the polymer, resin, pharmaceutical or
CC rubber industry. The present sequence encodes a naphthalene dioxygenase
CC (NDO) mutant alpha subunit, given in the exemplification of the present
CC invention
XX
SQ Sequence 4355 BP; 1065 A; 1143 C; 1171 G; 976 T; 0 U; 0 Other;
Query Match 5.0%; Score 49.8; DB 3; Length 4355;
Best Local Similarity 50.1%; Pred. No. 4.5e-05;
Matches 182; Conservative 0; Mismatches 172; Indels 9; Gaps 2;
QY 319 GTTGTGAGTCCGGCGCTAACTCATGACCTGCTCGTGGGATTAGCAGTGGG 378
1159 GTGGTGGCGGTGAGTGGCGGCACTACGATATCCGTCGCTACGCTCGCTCGTAA 1218
379 CCAGCAAAATTCCTCCCGGCCAGTTCGCTAGTAGAGCGCAGAGTGCAGCGGTG 438
1219 CCCTTCGAGTTCACCCGGACAGTAGCGACATT-----GCAGTTCAGTCTGAGCAT 1272
439 GTTCGGGCATATTCAATGGGAAATTTAAAGAACCCCGAAGGATATGGAGTTCATATT 498
1273 GCOCGTCCGTTTCAATGGCGAGTCTGCCAGATGACCAAGAAAT---GGAGTTCACATA 1329
499 AAGAGGGTACCCACAGACGATTTAGTCTTCGCTTTTCGAAATAGAAAGAGCGCT 558
1330 CGCAAGGTGCGGTGGCGGCTAACGGAGTAGTGTTCGACGACGTCGCGAAGTACA 1389
559 CGTCTATTGACGGGACCAATGGGCACATCTTCTTCGTCAGGGACCGCGCGAAG 618
1390 AGCATCAAGTTGAGCGGCGCACTTGGTAGCGCTTATTTCGTCAGAACACACCGCGCG 1449
619 AGCTTTGATTTGGCGGCGGCTGCTGATGCGCGCGCTATTTCAGCGCGCTCG 678
1450 ATGCTCTGTGTGGCGGTGGACCGGACTAGCACCGGTGCTGTCGATTGTCGCGCGCG 1509
679 ATG 681

DB 1510 CTG 1512
RESULT 8
ID AAA65346 standard; DNA; 14462 BP.
XX AC AAA65346;
XX 09-NOV-2000 (first entry)
DT NDO related complex alpha subunit nucleotide sequence SEQ ID NO:9.
XX Pseudomonas sp. strain NCIB 9816-4; naphthalene dioxygenase; NDO;
XX inducible multi-component enzyme; alpha subunit; mutant; chiral diol;
XX polymer; resin; pharmaceutical; rubber industry; bioremediation; ds.
XX Pseudomonas sp.
OS Synthetic.
OS WO200037480-A1.
XX 29-JUN-2000.
PF 26-OCT-1999; 99WO-US025079.
XX 26-OCT-1998; 98US-0105575P.
PR (IOWA) UNIV IOWA RES FOUND.
XX Parales R, Gibson D, Resnick S, Lee K;
PI WPI; 2000-452174/39.
DR P-PSDB; AAB12572.
XX Novel naphthalene dioxygenase mutant having a specific amino acid
PT substitution for preparing chiral diols for use in the polymer, resin,
PT pharmaceutical or rubber industry and for carrying out bioremediation.
XX Disclosure; Page 75-81; 151pp; English.
PS
XX The present invention describes a naphthalene dioxygenase (NDO) or NDO
CC related complex (I) comprising several polypeptides which contain an
CC alpha subunit that contains substituted amino acids at specific
CC positions. The polypeptides and host cells are useful for preparing (-)-
CC (1S,2R)-cis-naphthalene dihydrodiol, (- or +)-cis-biphenyl-3,4-
CC dihydrodiol, (1S,2R)-cis-phenanthrene-1,2-dihydrodiol, cis-1,2-
CC dihydroxyindan, 1,2-dihydroxy-1,2,3,4-tetrahydronaphthalene, which
CC involves contacting them with naphthalene, biphenyl, phenanthrene,
CC indene, and 1,2-dihydronaphthalene respectively. The polypeptides and the
CC host cells are also useful for preparing 1,2-dihydroxy-1,2-
CC dihydrophenanthrene or 3,4-dihydroxy-3,4-dihydrophenanthrene which
CC involves contacting them with phenanthrene, biphenyl, phenanthrene,
CC anthrane, phenanthrene, dibenzo(1,4)dioxan, acenaphthylene, 9,10-
CC dihydroanthracene, or 9,10-dihydrophenanthrene a corresponding dihydro
CC dihydroxy compound. The polypeptides and the host cells are useful for
CC preparing chiral diols for use in the polymer, resin, pharmaceutical or
CC rubber industry. The present sequence encodes a naphthalene dioxygenase
CC (NDO) mutant alpha subunit, given in the exemplification of the present
CC invention
XX
SQ Sequence 14462 BP; 3426 A; 3749 C; 4018 G; 3259 T; 0 U; 0 Other;
Query Match 4.6%; Score 45.6; DB 3; Length 14462;
Best Local Similarity 52.0%; Pred. No. 0.0022;
Matches 102; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 486 GGAGTTCATATTAAGAGGTTACCCACAGGACGATTAGTCTTCGTTTCGAAATAG 545

Db 4235 GGAGTCCACATACGACGAGTGGCGGTGGCGGTAACTGAGTATGTTTTCGAGCAGCT 4294
 Qy 546 AAAAGAGCGCTGCTATTTTGAAGGACCAATGGGCACATCTTTCTTCGTCAGG 605
 Db 4295 CCGGAGGTACAGCATCAATTTGAGCGGCCACCTGGTACGGCTATTTACGTCAGAA 4354
 Qy 606 GACCGGCGAAGAGTCTTTTGCATTTGGCGCGGTGGCGGCTCTCGTATGCGGCCGCTAT 665
 Db 4355 CCACACCGGCGGATGCTCTGTGTGGCGGTGAACCGGACTAGCACCAGTGTGTCGAT 4414
 Qy 666 TGCACGCGCTCGATG 681
 Db 4415 TATTCGGCGCGGCTG 4430

RESULT 9

ADA71938/c

ID ADA71938 standard; DNA; 2000 BP.

XX AC ADA71938;

XX DT 20-NOV-2003 (first entry)

XX DE Rice gene, SEQ ID 5263.

XX XX Plant; bacterial infection; fungal infection; viral infection; rice;

XX XX gene; ds.

XX OS Oryza sativa.

XX PN WO2003000898-A1.

XX PD 03-JAN-2003.

XX PF 22-JUN-2001; 2001WO-IB001105.

XX PR 22-JUN-2001; 2001WO-IB001105.

XX XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX XX WPI; 2003-175290/17.

XX PT Identifying at least one gene involved in plant resistance or response to

XX PT pathogenic infection for conferring resistance or tolerance to a plant to

XX PT bacterial, fungal or viral infection by determining or detecting plant

XX PT gene expression.

XX PS Claim 27; SEQ ID NO 5263; 899pp; English.

XX CC The present invention relates to a method (M1) for identifying genes

XX CC involved in plant resistance or response to pathogenic infection. M1

XX CC comprises identifying a gene whose expression is significantly altered in

XX CC the incompatible interaction of plant gene expression relative to

XX CC expression of the gene in an uninfected plant, in a mutant plant that

XX CC does not express a gene associated with response to pathogenic infection,

XX CC or in a corresponding incompatible or compatible interaction. (M1) is

XX CC useful for conferring resistance to resistance or tolerance to a plant to

XX CC bacterial, fungal or viral infection. The present sequence was used to

XX CC illustrate the invention.

XX SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match 4.3%; Score 43; DB 7; Length 2000;

Best Local Similarity 6.9%; Pred.No. 0.0054;

Matches 49; Conservative 324; Mismatches 334; Indels 0; Gaps 0;

Qy 270 GCAGGACAGTACGTCCTCCAGCATTCGAATCTCAAGATGGAAGGAGTGTGAGGT 329

Db 777 KSRVWKWKYATRYWYKQWAWTWWSWRKWSYMSGWRWSAARYCSRMKCAKTK 718

Qy 330 CCGGGCGCTAACTCATGACCTGCTGTCGTCGATTAAGCACTGATGGCGCAGCAATTT 389
 Db 717 YASSARWTKRAKRSYRVRRRWYKRKGWYRYYRWSRMTTRARMSRKRKRWAGASMKSCW 658
 Qy 390 COTCCCGCGCAGTTCTCCCTAGTAGAGCAGAGCAGTTGCGCAGCGTGGTTCGCGCATA 449
 Db 657 MYWRGARSMTYSKYSKCSAKCKKRYMTSSYVSTGMYSSYKSMSTSKVSMYMGKMT 598
 Qy 450 TTCAATGCGCAATTAAGAACCAGGAGGATATGGAGTCTATATTAAGAGGTACC 509
 Db 597 TWYTSKMGSTRSRKMGWSGRMYRMWVKORRYKRYMYKWKCTWRRCMYCYRWGYTMY 538
 Qy 510 CACAGCAGATTAGTCTGCTTTTCGAAATAGAAAAGAGCGCTCGTCTATTTT 569
 Db 537 TTSRSRMYTGRYKARYTSKRYMYKYKRYCWTYYYGYMYKCSYMYGYCKACKKCY 478
 Qy 570 GACGGGACCAATGGGCACATCTTTTCCTCGTCAGGAGACCGCGCGAAGAGTCTTTGCAT 629
 Db 477 AMCWKAAYSGMMYWRYSKYMNRMTKYMMSMYKKCRSMKYGAKGYCGKMMYTCYSY 418
 Qy 630 TGGCGCGGTGCGGGGCTCTCGTATGCGGCGCTATTGACGCGCTCGATGCGGGAAC 689
 Db 417 YMKWTYNGSYKYSRCYKYMRYMYKGYMYMYKMYMYKMYMYKMYMYKMYMYKMYMY 358
 Qy 690 AGACAAGCGGTAAAGTTGTTTCTACGGCTCAAGAACTCCGCGCGACGCTGTTGCGTGAAT 749
 Db 357 MSWYKSYKKYCTWYCYMKCVRKMYRKRKMYKMYKMYKMYKMYKMYKMYKMYKMY 298
 Qy 750 CGATATCGACATCGATGAGGACAGCTTGAAGTCTGTCAGCGCTCGATGCGGGAAC 809
 Db 297 YSRSMRTAGKWKRSWSRWSRYSWTYKTKMKWKSYMYSGWARSSTWSRSAKRT 238
 Qy 810 TAGCCTTTGGCAAGGCGCCACTGTTTATTATCATCAGTTGTCGACGCGCTGCTTGA 869
 Db 237 YKGYSTRRAKWRACRMYSAKRYRSTSYCGCSYCGSKWKYMSKSCSRMTCSWCS 178
 Qy 870 AACCTACCGGATACGAAATTTATCTTCGGTCCACCGCTATGTCGACGCTACTGT 929
 Db 177 CCYTCYGAACGSCCMYMGSCGCTRGWKRWSKYSMCKCKYSCCTKCYSGTYTRY 118
 Qy 930 CGGTATGCTGTCGCAAGGCTGTTTCCACGGGATCAAAATTCATTTTG 976
 Db 117 CKWYKSYKCYCYCYWYMSYMYRMYMKCMCSRSKCSWVSCAYCSTS 71

RESULT 10

ADA71938

ID ADA71938 standard; DNA; 2000 BP.

XX AC ADA71938;

XX DT 20-NOV-2003 (first entry)

XX DE Rice gene, SEQ ID 5263.

XX XX Plant; bacterial infection; fungal infection; viral infection; rice;

XX XX gene; ds.

XX OS Oryza sativa.

XX PN WO2003000898-A1.

XX PD 03-JAN-2003.

XX PF 22-JUN-2001; 2001WO-IB001105.

XX PR 22-JUN-2001; 2001WO-IB001105.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

DT 26-MAR-2002 (first entry)
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4432.
DE
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4432.
KW Drosophila melanogaster genomic polynucleotide SEQ ID NO 4432.
KW Drosophila melanogaster genomic polynucleotide SEQ ID NO 4432.
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4432.
OS Drosophila melanogaster genomic polynucleotide SEQ ID NO 4432.
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4432.
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4432.
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 4432; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 7785 BP; 2166 A; 2186 C; 2007 G; 1426 T; 0 U; 0 Other;
SQ
Query Match 3.9%; Score 38.4; DB 4; Length 7785;
Best Local Similarity 49.0%; Pred. No. 0.4;
Matches 102; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 127 TGCAAAATTCGAGTTACTTCGAGGGAATGTCCTCAATCAATGTCGCGGATGCTCCAGGACTT 186
DB 7243 TGCAAAATTCGAGTTACTTCGAGGGAATGTCCTCAATCAATGTCGCGGATGCTCCAGGACTT 7302
QY 187 TCTTCGCGAGATCGTGAGAGGGCAACCGGCATCTTGCATGCCAGTCCGCTCTCA 246
DB 7303 GTGTCCACATCTGTGCCAAGCTCGGTCTCCATGCCCATCATCGCGCAACTGCGCTTTC 7362
QY 247 GACCTGCGGATCAAAAGTCGAGTGCAGACAGTACGTCCTCAAGATTCCTCAAGA 306
DB 7363 TACTTGAGGGATAAAATCCCTTTCAGTTGCGAGTTCTGCTCAAGGAGCAAAAGTCAAA 7422
QY 307 ATGGAAGCGGAGTTGTTGAGTCCGGG 334
DB 7423 TTCGACACCGAACCCTATCGAGATCAAG 7450
RESULT 13
ABL17652/c
ID ABL17652 standard; DNA, 12118 BP.
XX
XX ABL17652;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4429.
DE

XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4429.
KW Drosophila melanogaster genomic polynucleotide SEQ ID NO 4429.
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4429.
OS Drosophila melanogaster genomic polynucleotide SEQ ID NO 4429.
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4429.
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 4429.
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 4429; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 12118 BP; 2692 A; 2753 C; 3140 G; 3533 T; 0 U; 0 Other;
SQ
Query Match 3.9%; Score 38.4; DB 4; Length 12118;
Best Local Similarity 49.0%; Pred. No. 0.51;
Matches 102; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 127 TGCAAAATTCGAGTTACTTCGAGGGAATGTCCTCAATCAATGTCGCGGATGCTCCAGGACTT 186
DB 1612 TGCAAAATTCGAGTTACTTCGAGGGAATGTCCTCAATCAATGTCGCGGATGCTCCAGGACTT 1553
QY 187 TCTTCGCGAGATCGTGAGAGGGCAACCGGCATCTTGCATGCCAGTCCGCTCTCA 246
DB 1552 GTGTCCACATCTGTGCCAAGCTCGGTCTCCATGCCCATCATCGCGCAACTGCTTTC 1493
QY 247 GACCTGCGGATCAAAAGTCGAGTGCAGACAGTACGTCCTCAAGATTCCTCAAGA 306
DB 1492 TACCTGAGGGATAAAATCCCTTTCAGTTGCGAGTTCTGCTCAAGGAGCAAAAGTCAAA 1433
QY 307 ATGGAAGCGGAGTTGTTGAGTCCGGG 334
DB 1432 TTCGACACCGAACCCTATCGAGATCAAG 1405
RESULT 14
ABL15161/c
ID ABL15161 standard; cDNA, 680 BP.
XX
XX ABL15161;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 39965.
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 39965.
KW Drosophila melanogaster expressed polynucleotide SEQ ID NO 39965.
KW Drosophila melanogaster expressed polynucleotide SEQ ID NO 39965.


```
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX XX
XX PF 23-MAR-2001; 2001WO-US009231.
XX XX
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PD
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-656860/75.
XX DR P-PSDB; ABB71058.
XX XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 39965; 21pp + Sequence Listing; English.
XX XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 680 BP; 193 A; 163 C; 164 G; 160 T; 0 U; 0 Other;

Query Match          3.8%; Score 37.6; DB 4; Length 680;
Best Local Similarity 54.3%; Pred. No. 0.19;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATCGATGAGGACACAGCTTGAGTTCGTTCCAGGCGATTACGGAACACACGGAT 810
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 GATCTCCGAGTCGATGATGATTGCTTCAGTTCTTCCATCAGCGACTGCTGGACGAAGGC 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 811 AGCCTTTGGCAAGGCCCACTGGTTTATTTCATCAGTTGTGCGACGCGCGCTGCTTGAA 870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 CTCCITGCGGATCATGGTGTCTCTTGTAGTTGGAGTTGTGGCGTACCGCAGCTTGCC 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 871 ACCCTACCGGAATACGAAAT 890
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 ATCCGCGCGGAACTCGAACT 160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
ABL15160/C
ID ABL15160 standard; cDNA; 2738 BP.
XX AC
XX AC ABL15160;
XX XX
XX DT 26-MAR-2002 (first entry)
XX XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 39962.
XX XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX XX pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX XX
XX PF 23-MAR-2001; 2001WO-US009231.
XX XX
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PD
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-656860/75.
XX DR P-PSDB; ABB71058.
XX XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 39962; 21pp + Sequence Listing; English.
XX XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX CC ABB72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2738 BP; 738 A; 643 C; 659 G; 698 T; 0 U; 0 Other;

Query Match          3.8%; Score 37.6; DB 4; Length 2738;
Best Local Similarity 54.3%; Pred. No. 0.41;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATCGATGAGGACACAGCTTGAGTTCGTTCCAGGCGATTACGGAACACACGGAT 810
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1299 GATCTCCGAGTCGATGATGATTGCTTCAGTTCTTCCATCAGCGACTGCTGGACGAAGGC 1240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 811 AGCCTTTGGCAAGGCCCACTGGTTTATTTCATCAGTTGTGCGACGCGCGCTGCTTGAA 870
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1239 CTCTTCCGATCATGGTGTCTCTTGTAGTTGGAGTTGTGGCGTACCGCAGCTTGCC 1180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 871 ACCCTACCGGAATACGAAAT 890
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1179 ATCCGCGCGGAACTCGAACT 1160
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: June 27, 2004, 05:02:43
Job time : 435.15 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 04:30:32 ; Search time 2839.31 Seconds
(without alignments)
10412.231 Million cell updates/sec

Title: US-10-658-691-3
Perfect score: 990
Sequence: 1 agtaccacactcaaaattga.....attttgaagcatttttctaa 990

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pla:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_lam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vri:*
28: gb_gss1:*
29: gb_gss2:*

5	37.8	3.8	440	9	AA813209	AA813209	aj32e04.s
6	37.8	3.8	552	28	CC047819	CC047819	3591.1.98
7	37.8	3.8	634	12	BI367122	BI367122	RE52813.5
c 8	37.8	3.8	1176	28	BZ551949	BZ551949	pacsl-60
c 9	37.6	3.8	369	12	BI167687	BI167687	RE08123.5
c 10	37.6	3.8	441	11	AY231918	AY231918	Drosophila
c 11	37.6	3.8	441	11	AY232121	AY232121	Drosophila
c 12	37.6	3.8	517	10	BE978534	BE978534	be79a04.y
c 13	37.6	3.8	541	12	BI167742	BI167742	RE08194.5
c 14	37.6	3.8	557	12	BI483125	BI483125	RE65966.5
c 15	37.6	3.8	606	12	BI167128	BI167128	RE07446.5
c 16	37.6	3.8	645	12	BI356036	BI356036	RE42061.5
c 17	37.6	3.8	661	12	BI236552	BI236552	RE32623.5
c 18	37.6	3.8	661	12	BI367637	BI367637	RE53440.5
c 19	37.6	3.8	661	12	BI371843	BI371843	RE58850.5
c 20	37.6	3.8	663	12	BI230435	BI230435	RE14116.5
c 21	37.6	3.8	663	12	BI242569	BI242569	RE39946.5
c 22	37.6	3.8	675	13	BQ705266	BQ705266	Y1A10A04
c 23	37.6	3.8	677	12	BI632106	BI632106	SD25703.5
c 24	37.6	3.8	683	12	BI228632	BI228632	RE26236.5
c 25	37.6	3.8	684	12	BI230656	BI230656	RE14375.5
c 26	37.6	3.8	686	12	BI485460	BI485460	RE68784.5
c 27	37.6	3.8	706	12	BI371383	BI371383	RE58261.5
c 28	37.6	3.8	871	13	BQ705783	BQ705783	Y1B04F09
c 29	37.2	3.7	1111	29	CNS01497	CNS01497	Drosophila
c 30	36.8	3.7	884	14	CD299445	CD299445	AGENCOURT
c 31	36.8	3.7	887	12	BI834500	BI834500	603084671
c 32	36.8	3.7	932	13	BUI13845	BUI13845	603130341
c 33	36.6	3.7	688	13	BW261927	BW261927	BW261927
c 34	36.2	3.7	657	28	BH507017	BH507017	BOHQD45TR
c 35	36.2	3.7	702	28	BH936230	BH936230	odf10f02
c 36	36.2	3.7	712	28	BH513796	BH513796	BOGUX75TR
c 37	36.2	3.7	725	28	BH730081	BH730081	BOMEE23TF
c 38	36	3.6	434	9	AL817274	AL817274	AL817274
c 39	35.6	3.6	1087	9	AL575881	AL575881	AL575881
c 40	35.4	3.6	445	10	BE495071	BE495071	WHE1267.B
c 41	35.4	3.6	577	12	BQ082775	BQ082775	fu25906.Y
c 42	35.4	3.6	633	13	BQ263855	BQ263855	faa18h04
c 43	35.4	3.6	649	28	BZ852591	BZ852591	CP240.219
c 44	35.4	3.6	655	12	BI891291	BI891291	ZF637-3.0
c 45	35.4	3.6	658	12	BM184423	BM184423	fv68b06.Y

ALIGNMENTS

RESULT 1
BX381961
LOCUS BX381961 1201 bp mRNA linear EST 08-MAY-2003
DEFINITION BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA
clone CS0D1072YF05 3-PRIME, mRNA sequence.
ACCESSION BX381961
VERSION BX381961.1 GI:30453007
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS Li.W.B., Gruber.C., Jesse.J. and Polayes.D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contract: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D1072CC03N1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"

FEATURES
source

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	42.8	4.3	1201	13	BX381961	BX381961
c 2	40.8	4.1	600	12	BI229988	BI229988
c 3	38.2	3.9	615	12	BJ260819	BJ260819
c 4	37.8	3.8	291	28	BH569723	BH569723

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

[illegible]

RESULT 2	LOCUS	DEFINITION	600 bp	mRNA	linear	EST 11-JUL-2001
BI229988	BI229988/c	RE27502.5prime RE Drosophila melanogaster normalized Embryo pFlc-1				
		Drosophila melanogaster cDNA clone RE27902.5 similar to mago:				
		FBan0003401 located on: 2R 57C2-57C2.1: 04/12/2001, mRNA sequence.				
BI229988						
BI229988						
BI229988.1		GI:14697252				
EST.						
Drosophila melanogaster		(fruit fly)				
Drosophila melanogaster						
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;						
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;						
ACCESION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						

[illegible]

National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source

1. .615
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whh27all"
/tissue_type="spike at heading date"
/dev_stage="Feekes scale 10.5"
/clone_lib="Y. Ogihara unpublished cdna library, wh_h"

ORIGIN

Query Match 3.9%; Score 38.2; DB 12; Length 615;
Best Local Similarity 49.7%; Pred. No. 9.4;
Matches 97; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
QY 144 CGAAGGATGTCATCAATGTCGGGATGTCCTCAGGACTTCTTCGCGGATCGTGA 203
DB 471 CGACGTGGAGGTAGAACAGAGCGGTCCATGCGCGGAGTAGCGCGCCAGGATGGCGC 412
QY 204 GAAGGCAACCGGCATCTTCATGTCAGTGCCTGCTCAGACTCGGATCAAGCT 263
DB 411 CGATGGCATGAGATGCCCGAGAGGTTGAGCCCTCGTGGCCCGGAGGC 352
QY 264 CGCAGTGCAGGACAAGTACGTCCTCCACGATTCCTCAAGATCGAAGCGGAGTTCT 323
DB 351 CGTAGGGAGAGGCGCCGTCGCGGAGTCTTCCAGTGTGAGTGTGAAGTGAAGGTGAAGGAGG 292
QY 324 TGAGGTCCGGCGCT 338
DB 291 TCTGTGCTGCTGCT 277

RESULT 4
BH569723/c
LOCUS
DEFINITION BOHQ45R BOHQ Brassica oleracea genomic clone BOHQ45, genomic survey sequence.
ACCESSION BH569723
VERSION BH569723.1 GI:17821562
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 291)
Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
CONTACT: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.
Location/Qualifiers
1. .291
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000H3"
/db_xref="taxon:3712"
/clone="BOHQ45"
/clone_lib="BOHQ45"
/note="vector: pHS01; Site 1: BstXI; 2-3 kb sheared

FEATURES
source

1. .291
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000H3"
/db_xref="taxon:3712"
/clone="BOHQ45"
/clone_lib="BOHQ45"
/note="vector: pHS01; Site 1: BstXI; 2-3 kb sheared

genomic DNA inserted into pHS01 using BstXI linkers"
Query Match 3.8%; Score 37.8; DB 28; Length 291;
Best Local Similarity 47.3%; Pred. No. 8.5; Indels 0; Gaps 0;
Matches 114; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 736 GCTGTTCGGTGGATCGATATCGACATCGATGAGGACAGCTTGAGTGGTCCAGGAGTT 795
DB 262 GCTGATTTAGGGGATGAACCTTGATGTGGATAAGCTCTTGTCTTGAGATAAGAACCCCAT 203
QY 796 ACGGAAGACAGGATAGCCTTTGGCAAGGCCCCCTGTTTATTCATCAGGTTGTGCAC 855
DB 202 GCTGAAACATATCATGAGATTTGAACGTTGTGAATTTATTTGATGAGAGATAGATT 143
QY 856 GCAGCGCTGCTGAAACCTTACCGGAATACGAAATTTATCTTCCGCTCCACCGCCTATG 915
DB 142 GACTACTTTGTTAAACCTCGTTTGAACCTGTAGTGTGTTTGTGTTCCACCTATTGCT 83
QY 916 GTCACGCTACTGTCCTGATGCTCTCGGCAAGGCTTCCACCGGATCAATTCATTT 975
DB 82 TTGATTCATGATCATCTGAACCTCAAGAAGCATTTATATAGTCATATCATATCTTTA 23
QY 976 G 976
DB 22 G 22

RESULT 5
AA813209

LOCUS
DEFINITION

ACCESSION
VERSION
KEYWORDS
SOURCE

ORGANISM
Homo sapiens (human)

REFERENCE
AUTHORS
TITLE

JOURNAL
COMMENT

AA813209
aj32e04.s1 Soares testis_NHT Homo sapiens cdna clone 1392030 3',
mRNA sequence.
AA813209
AA813209.1 GI:2883194
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 440)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyT not found
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 378.
Location/Qualifiers
1. .440
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="1392030"
/sex="male"
/lab host="DH10B"
/clone_lib="Soares testis NHT"
/note="Vector: pMT30-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cdna was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5]
TGTTACCAATCTGAAGTGGAGCGGCCCCCAATTTTTTTTTTTT 3'.

FEATURES
source

1. .440
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="1392030"
/sex="male"
/lab host="DH10B"
/clone_lib="Soares testis NHT"
/note="Vector: pMT30-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cdna was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5]
TGTTACCAATCTGAAGTGGAGCGGCCCCCAATTTTTTTTTTTT 3'.

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization to cots, and was constructed by Bento Soares and M. Fatima Bonaldo. "

ORIGIN

Query Match 3.8%; Score 37.8; DB 9; Length 440;
Best Local Similarity 47.6%; Pred. No. 10;
Matches 108; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

Qy 24 GAAGCGCCAGGACCTGGCTGAGGAGAGAGCTGTGTGCTCAGCAGCTTGTCTAATGG 83
|||
Db 64 GCAGCCCGAAGCCCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 123
|||
Qy 84 TATCGGATTCGTACGAGTGTCATCGGAGGTTTCCGAGGATGCAATTCGAGTTACT 143
|||
Db 124 TTTGAGTTATCTGTAGTGGAGACGACGAGAGCTGAACCGGACGGCCAGGG 183
|||
Qy 144 CGAAGGAATGTCATCAATATGCGCGGATCTCCAGGACTTTCTTCGCGAGATCGTGA 203
|||
Db 184 TGATCAGCAAGTTCACTCTGTTCTCTATGCGCTGCGCAGGCTGCGGAGGCGCGG 243
|||
Qy 204 GAAGGCAACCCCATCTTTCATGCGAGTGCCTGCTGCTCTCAGACC 250
|||
Db 244 GGTGAGGCGCGGATATGCGCAGGACGGGTTGAGCGCGCAGACC 290
|||

RESULT 6

LOCUS CC047819

DEFINITION 3591_1_98_1_B10.X.1 3591 - RescueMu Grid P Zea mays genomic, genomic survey sequence.

ACCESSION CC047819

VERSION CC047819.1 GI:29462710

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

REFERENCE 1 (bases 1 to 552)

AUTHORS Walbot,V.

TITLE Maize genomic sequences found using engineered RescueMu transposon

JOURNAL Unpublished (2001)

COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 3591_1_98_1 ROW: 5
Class: transposon-tagged.
Location/Qualifiers
1..552
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73/K55"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="3591 - RescueMu Grid P"
/note="Organ: leaf; Vector: RescueMu (engineered from Bluescript backbone); Site 1: BamHI; Site 2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid P was grown at Molokai in 2002. DNA was extracted from leaf strips, double digested using BamHI

FEATURES

source

1..552
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="mixed background W23/A188/B73/K55"
/db_xref="taxon:4577"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="3591 - RescueMu Grid P"
/note="Organ: leaf; Vector: RescueMu (engineered from Bluescript backbone); Site 1: BamHI; Site 2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid P was grown at Molokai in 2002. DNA was extracted from leaf strips, double digested using BamHI

and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

Query Match 3.8%; Score 37.8; DB 28; Length 552;
Best Local Similarity 57.0%; Pred. No. 12;
Matches 69; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 632 GCGCGGTGCGGGCTCTCGTATGCGCGCTTATTCACGCGCTCGATCGCGGAACAG 691
|||
Db 261 GCAGCGCGCGCTCTCGATCGCGGAGCTATTCTGCGCGTTCGGAGAGGTGAATC 320
|||
Qy 692 ACAAGCGGTAAAGTTGTTCTACGGCTCAAGAACTCCGCGGACGCTGTTTCGCTGGATCG 751
|||
Db 321 CTACCGCGTGCAGCTCACTTCTCATAGAACCACTTGGCGCATTCGATCATG 380
|||
Qy 752 A 752
|
Db 381 A 381
|

RESULT 7

LOCUS BI367122/c

DEFINITION RE52813.Sprime RE Drosophila melanogaster normalized Embryo pFLC-1 Drosophila melanogaster cDNA clone RE52813 5 similar to mego: FBan009401 located on: 2R 57C2-57C2; 05/14/2001, mRNA sequence.

ACCESSION BI367122

VERSION BI367122.1 GI:15063150

KEYWORDS EST.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

REFERENCE 1 (bases 1 to 634)

AUTHORS Scapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B., Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Friese,E., George,R., Gonzales,M., Guarin,H., Harris,N., Li,P., Liao,G., Misra,S., Mungall,C.J., Nuncio,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Celniker,S. and Rubin,G.M.

TITLE BDGP/HMI RE Drosophila EST Project

JOURNAL Unpublished (2001)

COMMENT Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
hit genomic AE003453: arm:2R [16010240,16312786]
estimated-cyto:57B20-57D11: 05/14/2001
Plate: RE.528 row: B column: 1
High quality sequence stop: 500.
Location/Qualifiers
1..634
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo pFLC-1"
/note="Organ: embryo; Vector: pFLC1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

FEATURES

source

1..634
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha Tona"
/clone_lib="RE Drosophila melanogaster normalized Embryo pFLC-1"
/note="Organ: embryo; Vector: pFLC1; Site 1: XhoI; Site 2: BamHI; Library was kindly generated by Piero Carninci at the RIKEN. The library was normalized and excised using Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 3.8%; Score 37.8; DB 12; Length 634;

Best Local Similarity 51.5%; Pred. No. 13;
Matches 87; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 722 GAATCCGCGCGAGCTGTTCGGTGCATCGATATCGATAGGACACAGCTTGAGG 781
Db 353 GAGCTCAGGCGAGATCGTCTCTCGATGATCTCCGATCGATGATGATTCCTCAT 294

QY 782 TCCTCCAGGCGATTACGGAAGACACCGATAGCTTTGGCAAGGCCCATCGTTTATTC 841
Db 293 TCCTCCATCAGGACTGGTGCAGAGGCGCTCTTCGGATCATGCTGCTCTGTAG 234

QY 842 ATCAGGTTCGAGCGAGCGCTCTTCAACCTACCGNATACGAAAT 890
Db 223 TTGGAGTGTGGCGTACCGCAGCTTGCCATCCGCGCGAACTCGAACT 185

RESULT 8
BZ551949/c 1176 bp DNA linear GSS 17-DEC-2002
LOCUS
DEFINITION
pacsi-60_3407.xl pacsi-60 Pseudomonas aeruginosa genomic clone
pacsi-60_3407, genomic survey sequence.

ACCESSION
VERSION BZ551949
KEYWORDS GSS.
SOURCE
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE
1 (bases 1 to 1176)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) in press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: ckraymond@u.washington.edu
Class: shotgun.

FEATURES
source
1. .1176
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"
/db_xref="taxon:287"
/clone="pacsi-60_3407"
/clone_lib="pacsi-60"
/note="clinical isolate 1-60 Whole genomic shotgun
library."

ORIGIN
Query Match 3.8%; Score 37.8; DB 28; Length 1176;
Best Local Similarity 58.4%; Pred. No. 17;
Matches 66; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 389 TCCTCCCGCGCGAGTTCCTGCTAGTAGAGGAGAGGATGCGAGCGGTGTCGCGCAT 448
Db 198 TCGCAACGCGCGGTCCACTGTGTCGACGAGAGCATGTCGCGGCGTGGCGCGGAC 139

QY 449 ATTCAATGCGCAATTAAAGAACCCCGAGGCATATGGAGTCTTATTAAG 501
Db 138 AGTTGCGCGTTCATCTGCGCAATGGAAGGCGTAGAAGGAATTCGATCAAG 86

RESULT 9
B1167687/c 369 bp mRNA linear EST 09-JUL-2001
LOCUS
DEFINITION
RE08123.5prime RE Drosophila melanogaster normalized Embryo pf1c-1
Drosophila melanogaster cDNA clone RE08123 5 similar to mago;
FBan0009401 located on: 2R 57C2-57C2; 04/11/2001, mRNA sequence.

B1167687
B1167687.1 GI:14633494
EST.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 369)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
Mirza, S., Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celniker, S. and
Rubin, G.M.
BDGP/HIMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu>
hit genomic AE003453: arm:2R [16010240..16312786]
estimated-cyto:57B20-57D11: 04/11/2001
Plate: RE.81 row: B column: 11
High quality sequence stop: 315.
Location/Qualifiers
1. .369
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE08123"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DHS-alpha Iona"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pf1c-1"
/note="Organ: embryo; Vector: pFLcl; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN
Query Match 3.8%; Score 37.6; DB 12; Length 369;
Best Local Similarity 54.3%; Pred. No. 11;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATCGATGAGGACAGCTTGAGTTCGTCAGGAGTACGGAACACGAT 810
Db 322 GATCTCCGAGTCGATGATTCGCTTCAGTTCTTCCATCAGGACTGTTGACGAAGC 263

QY 811 AGCCTTTGGCAAGGCGCCACTGTTTATTCATCAGTTTGTGACGCGCTGCTTGA 870
Db 262 CTCCTTGGGATCATGGTGTCTTCTGTAGTTGAGTTGTCGCTACCGAGCTTGC 203

QY 871 ACCCTACGGAATACGAAT 890
Db 202 ATCCGCGCGAACTCGAACT 183

RESULT 10
AY231918/c 441 bp mRNA linear HTC 04-NOV-2003
LOCUS
DEFINITION
Drosophila yakuba clone yak-em_mago mRNA sequence.
AY231918
ACCESSION
VERSION AY231918.1 GI:38048076
KEYWORDS HTC.
SOURCE
ORGANISM Drosophila yakuba
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 441)
AUTHORS Domazet-Lozo, T. and Tautz, D.
TITLE An evolutionary analysis of orphan genes in Drosophila
JOURNAL Genome Res. 13 (10), 2213-2219 (2003)
MEDLINE 22887302
PUBMED 14525923
REFERENCE 2 (bases 1 to 441)
AUTHORS Domazet-Lozo, T. and Tautz, D.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2003) Evolution Genetics, Institute for Genetics,
Weyertal 121, Cologne 50931, Germany
FEATURES
source Location/Qualifiers
1..441
/organism="Drosophila yakuba"
/mol_type="mRNA"
/db_xref="taxon:7245"
/clone="yak-em_mago"
<1..>441
/codon_start=1
/product="similar to Drosophila melanogaster mago"
/protein_id="AK09941.1"
/db_xref="GI:38048077"
/translation="MSTEDFYLRYYVGHKFGHEFLFEFRPDGKLRVANNVYKND
TMRKEAFVHQSVMEELKRIIDSEIMQEDDLPPPPDRVGRQLEIVIGDEHISFTT
SKTGLVDVNRKDPGLRCFYLVQDLKCLVSLGLHFKIKPI"

CDS
Query Match 3.8%; Score 37.6; DB 11; Length 441;
Best Local Similarity 54.3%; Pred. No. 12; Mismatches 64; Indels 0; Gaps 0;
Matches 76; Conservative 0

ORIGIN
1 517 bp mRNA linear EST 04-NOV-2003
LOCUS AY232121/c
DEFINITION Drosophila yakuba clone yak-ad_mago mRNA sequence.
ACCESSION AY232121
VERSION AY232121.1 GI:38048482
KEYWORDS HTC.
SOURCE Drosophila yakuba
ORGANISM Drosophila yakuba
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 441)
Domazet-Lozo, T. and Tautz, D.
An evolutionary analysis of orphan genes in Drosophila
Genome Res. 13 (10), 2213-2219 (2003)
MEDLINE 22887302
PUBMED 14525923
REFERENCE 2 (bases 1 to 441)
AUTHORS Domazet-Lozo, T. and Tautz, D.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-2003) Evolution Genetics, Institute for Genetics,
Weyertal 121, Cologne 50931, Germany
FEATURES
source Location/Qualifiers
1..441
/organism="Drosophila yakuba"
/mol_type="mRNA"
/db_xref="taxon:7245"
/clone="yak-ad_mago"

CDS
Query Match 3.8%; Score 37.6; DB 11; Length 441;
Best Local Similarity 54.3%; Pred. No. 12; Mismatches 64; Indels 0; Gaps 0;
Matches 76; Conservative 0

ORIGIN
1 517 bp mRNA linear EST 04-OCT-2000
LOCUS BE978534/c
DEFINITION Drosophila melanogaster adult testis library Drosophila
melanogaster cDNA clone bs79a04 5', mRNA sequence.
ACCESSION BE978534
VERSION BE978534.1 GI:10609959
KEYWORDS EST.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 517)
Andrews, J., Bouffard, G. and Oliver, B.
Drosophila melanogaster testis expressed sequence tags
Unpublished (1999)
Contact: Brian Oliver
Laboratory of Cellular and Developmental Biology
NIDDK, National Institutes of Health
6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
Fax: (301) 496 5239
Email: oliver@helix.nih.gov,
http://www.niddk.nih.gov/intram/people/boliver.htm
Tissue isolation and library construction performed at the National
Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
http://www.niddk.nih.gov/intram/people/boliver.htm). DNA sequencing
and analyses performed by National Institutes of Health Intramural
Sequencing Center (NISC; see http://www.nisc.nih.gov).
Plate: 79 row: a column: 04
Seq primer: M13RP1 reverse primer (ABI).
Location/Qualifiers
1..517
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/strain="y[*] w[67c1]/y"
/db_xref="taxon:7227"
/clone="bs79a04"
/sex="male"
/dev_stage="1-5 day adult"
/lab_host="SOLR (Stratagene)"
/clone_lib="Drosophila melanogaster adult testis library"
/note="Organ: testis; Vector: pBlueScript SK (Stratagene);
Site_1: EcoR I; Site_2: Xho I; Testes dissected from 1-5
day adult y[*] w[67c1]/y males raised at 25°C. RNA
isolated using Trizol (Life Technologies) and a single

Query Match 3.8%; Score 37.6; DB 12; Length 557;
Best Local Similarity 54.3%; Pred. No. 13;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATCGATGAGGACAGCTTGAGTTCGTCAGGACAGTACGGAACACCGGAT 810
|||||
Db 320 GATCTCGAGTCGATGATGATTCGCTTCAGTTCCTCCATCAGGACTGTGGACGAGGC 261
|||||

QY 811 AGCCTTTGGCAAGGGCCCACTGTTTTATTCATCAGGTTGTGACGACGAGCGCTGCTTGAA 870
|||||

Db 260 CTCCTTTGGGATCATGCTGTGCTTCTTGTAGTTGGAGTTGTTGGCGTACCGCAGCTTGCC 201
|||||

QY 871 ACCCTACCGGAATACGAAAT 890
|||||

Db 200 ATCCGGCCGGAACCTCGAACT 181
|||||

RESULT 15
BI167128/c
LOCUS
DEFINITION
RE07446 5prime RE Drosophila melanogaster normalized Embryo pFlc-1
Drosophila melanogaster CDNA clone RE07446 5 similar to mago;
Fban0009401 located on: 2R 57C2-57C2; 04/11/2001, mRNA sequence.

BI167128
BI167128.1 GI:14632935
EST.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 606)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B.,
Carlson, J., Champe, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E.,
George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G.,
Misra, S., Mungall, C. J., Nuno, D., Pacieb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S. E., Celnik, S. and
Rubin, G. M.

TITLE
BDGP/HMI RE Drosophila EST Project
JOURNAL
Unpublished (2001)
COMMENT
Contact: Stapleton, M.
BDGP

Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
hit genomic AE003453: arm:2R [16010240,16312786]
estimated-cyto:57B20-57D11: 04/11/2001
plate: RE.74 row: D column: 10
High quality sequence stop: 504.
Location/Qualifiers
1. .606
/organism="Drosophila melanogaster"
/mol_type="mRNA"
/db_xref="taxon:7227"
/clone="RE07446"
/sex="male and female"
/dev_stage="0-24 hours mixed stage embryonic"
/lab_host="DH5-alpha tonA"
/clone_lib="RE Drosophila melanogaster normalized Embryo
pFlc-1"
/notes="Organ: embryo; Vector: pFlc1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

ORIGIN

Query Match 3.8%; Score 37.6; DB 12; Length 606;
Best Local Similarity 54.3%; Pred. No. 14;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATCGATGAGGACAGCTTGAGTTCGTCAGGACAGTACGGAACACCGGAT 810
|||||

Db 302 GATCTCGAGTCGATGATGATTCGCTTCAGTTCCTCCATCAGGACTGTGGACGAGGC 243
QY 811 AGCCTTTGGCAAGGGCCCACTGTTTTATTCATCAGGTTGTGACGACGAGCGCTGCTTGAA 870
|||||

Db 242 CTCCTTTGGGATCATGCTGTGCTTCTTGTAGTTGGAGTTGTTGGCGTACCGCAGCTTGCC 183
|||||

QY 871 ACCCTACCGGAATACGAAAT 890
|||||

Db 182 ATCCGGCCGGAACCTCGAACT 163
|||||

Search completed: June 27, 2004, 09:59:09
Job time : 2847.31 secs


```

; GENERAL INFORMATION:
; APPLICANT: Falb, Dean A.
; APPLICANT: Gimeno, Carlos J.
; TITLE OF INVENTION: Conservin Compositions and Therapeutic
; TITLE OF INVENTION: and Diagnostic Uses Therefor
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/688,609
; FILING DATE:
; PRIORITY APPLICATION DATA:
; PRIORITY APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: MIQ-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 444 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;
US-08-688-609-7

Query Match 3.8%; Score 37.6; DB 1; Length 444;
Best Local Similarity 54.3%; Pred. No. 0.0095;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 751 GATATCGACATCGATGAGGACAACTTGAGGTGCTCCAGGCGATTACGGAGACACGGAT 810
Db 210 GATCTCCGAGTCGATGATGATTCGCTTCAGTTCTTCATCAGCGACTGGTGGAGGAGGC 151
QY 811 AGCCTTTGGCAAGGGCCCACTGGTTTATTATTCATCAGGTTGTCGACGACGAGCTGCTTGAA 870
Db 150 CTCCTTCGGGATCATGTGTGCTTCTTGTAGTTGGAGTTGTTGGGTACCGCAGCTTGCC 91
QY 871 ACCCTACCGGAATACGAAAT 890
Db 90 ATCGGCCGGAACTCGAACT 71

RESULT 3
US-09-002-832-7/c
; Sequence 7, Application US/09002832
; Patent No. 6031076
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean A.
; APPLICANT: Gimeno, Carlos J.
; TITLE OF INVENTION: Conservin Compositions and Therapeutic
; TITLE OF INVENTION: and Diagnostic Uses Therefor
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA

```

APPLICATION NUMBER: JP Hei-6-179689
FILLING DATE: 08-JUL-1994
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6379 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
STRAIN: No. 561220Ardia corallina B-276 (FERM P-4094; FERM
STRAIN: BP-5124; ATCC 31338)
FEATURE:
NAME/KEY: CDS
LOCATION: 910..1935
OTHER INFORMATION: /product= "amoA"
FEATURE:
NAME/KEY: CDS
LOCATION: 1935..2285
OTHER INFORMATION: /product= "amoB"
FEATURE:
NAME/KEY: CDS
LOCATION: 2300..3802
OTHER INFORMATION: /product= "amoC"
FEATURE:
NAME/KEY: CDS
LOCATION: 3805..4830
OTHER INFORMATION: /product= "amoD"
US-08-499-215-1

Query Match 3.7%; Score 36.8; DB 1; Length 6379;
Best Local Similarity 58.0%; Pred. No. 0.1;
Matches 65; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 872 CCTACCGGAATACGAATTTATCTTCCGCTCCACCGCTTGTGTCGAGCTACTGTCC 931
DB 4691 CCGAATCGGCATGACGTCTACCTGTGGGGCGCCACCATGATCGACGCGCTGTGC 4750
QY 932 GTATGCTGTGCGCAAGGTTTCCACGCGATCAATTTGACGCAAT 983
DB 4751 CGCTGCTGTGCGAGGGGGTGGCCCGACGACATCTACTACGACGCAAT 4802

RESULT 5
US-09-453-956-1
Sequence 1, Application US/09453956
Patent No. 6472191
GENERAL INFORMATION:
APPLICANT: Yano, Tetsuya; No. 6472191oto, Tsuyoshi; Imamura, Takeshi;
APPLICANT: Canon Kabushiki Kaisha
TITLE OF INVENTION: DNA Fragment Carrying Toluene Monooxygenase Gene,
TITLE OF INVENTION: Recombinant Plasmid, Transformed Microorganism,
TITLE OF INVENTION: Method for Degrading Halogenated Aliphatic Hydrocarbon
TITLE OF INVENTION: Compounds and Aromatic Compounds, and
TITLE OF INVENTION: Method for Environmental Remediation
FILE REFERENCE: CF014074US
CURRENT APPLICATION NUMBER: US/09/453,956
CURRENT FILING DATE: 1999-12-03
EARLIER APPLICATION NUMBER: JP, P1998-344506
EARLIER FILING DATE: 1998-12-03
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 5331
TYPE: DNA
ORGANISM: Ralstonia eutropha
FEATURE:
NAME/KEY: CDS
LOCATION: (200)..
OTHER INFORMATION: tomK
NAME/KEY: CDS

LOCATION: (446)..
OTHER INFORMATION: tomL
FEATURE:
NAME/KEY: CDS
LOCATION: (1474)..
OTHER INFORMATION: tomM
FEATURE:
NAME/KEY: CDS
LOCATION: (1781)..
OTHER INFORMATION: tomN
FEATURE:
NAME/KEY: CDS
LOCATION: (3345)..
OTHER INFORMATION: tomO
FEATURE:
NAME/KEY: CDS
LOCATION: (3741)..
OTHER INFORMATION: tomP
FEATURE:
NAME/KEY: CDS
LOCATION: (4801)..
OTHER INFORMATION: tomQ
US-09-453-956-1
Query Match 3.6%; Score 35.8; DB 4; Length 5331;
Best Local Similarity 43.4%; Pred. No. 0.2;
Matches 279; Conservative 0; Mismatches 352; Indels 12; Gaps 2;
QY 48 AGGGAAGAGCCTGTGGTCTCAGCACCTTGAATCGATTCGATTCGGTACGAGTGTC 107
DB 3794 AGGCGAGACCATTCCTCGATCGGCGCTCGCCNAGGCATCTATATCCGCGCTGTG 3853
QY 108 ATCGGAGGTTGCGAGATGCAAAATTCAGTTACTCGAAGGATGTCATCAATGTC 167
DB 3854 CCACGCGCTGTGCGGAGCTTGCAGAGTCTCGGTCTCTGACGCGAGGCGCCTGGGCGA 3913
QY 168 GCCGATGCTCCAGGACTTTCTTCCGAGATCTGTGAGAAGGCAACCCCATCTTTCATG 227
DB 3914 GGCCTACCGCTGCGGTGATGAGATTTGAGCGCGAGAGGCGCTGGCGTGTG 3973
QY 228 CCAGTGCCTGCTCTCAGACCTCGGATCAAAATTCGAGTGCAGTGCAGAGCAAGTACGTCC 287
DB 3974 CGCGACGCTGCAGGCGGATACCCATCGAGGCCGATGTCCACAAAGAGCCCGGCGGCGA 4033
QY 288 AACGATTCATCTCAAGATGGAAGCGGAAGTTGTGAGGTCCGGGCGCTAACTCATGA 347
DB 4034 GATTATCCCGGTGCGGGATTTGAGGCGGACGTGATGTCATCGACAGCTCACCCGAC 4093
QY 348 CCGTGTGCTCGGTGCGATTACGCACTGATGGGCGCAGCAAAATTTCTTCCCGGCGAGTTCG 407
DB 4094 CATCAAGGCGATCCGCTGCGGCTCGCGAGCGGATGGTTTCCAGGCGGCGAGTACGT 4153
QY 408 CCTAGTAGGCGAGCAGTTCGAGGCGTGTTCGCGCATATTCAATGGCGAATTTAA 467
DB 4154 CCAGTTCGAG---ATCCCGGGCTTGGGCGAGACCCGCGCTTTCTCGATCGCAACGCGCC 4210
QY 468 GAACCCCGA-----AGGCATATGGGAGTTCTATATTAGAGGGTACCCACAGGAGC 518
DB 4211 GCGGAGCTGCGCGGACCGCGAGATCGAGCTGAACGTGCGGCGAGTCCCGGGCGGCT 4270
QY 519 ATTAGTCTTGGCTTTTCGAAAATAGAAAAGAGCGCTCGTCTATTTTTACCGGAC 578
DB 4271 TGACACCGCTACCTGACGAGCTCGCCCGCGGGATCGCGTGGCTTTGTCGGGAC 4330
QY 579 AATGGGCACATCTTTCTCCGTCAGAGGACCGCGCGAAGAGTCTTTGCAATGGGCGGCG 638
DB 4331 CTATGGCCGCTTCTTGTGCGCGCTCGCGCGGCTCGCGGATGATCTTCATGCGGCGG 4390
QY 639 TGCGGCGCTCTCGTATGCGGCGCTATTGACAGCGGCTCGATG 681
DB 4391 CTGCGGCGCTGCGAGCCCGCTCCATGATCTCGACCTGCTG 4433

RESULT 6

US-09-784-316-3
; Sequence 3, Application US/09784316
; Patent No. 6461843
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THREOF
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/784,316
; CURRENT FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 65042
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc feature
; LOCATION: (1)-(65042)
; OTHER INFORMATION: n = A,T,C or G
US-09-784-316-3

Query Match 3.5%; Score 34.8; DB 4; Length 65042;
Best Local Similarity 62.8%; Pred. No. 2.2;
Matches 54; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 534 TTTCGAAATAGAAAGAGCGCTCGTCTATTTTTCACGGGACCAATGGGCACATCTTT 593
DB 6894 TTTTGAATAGAAAGAGAGAGTCAACAGTAGAAGACTAGACCTCTGGGCAATACAAT 6953
QY 594 CTTCCGTCACAGGACCGCCGAAAGA 619
DB 6954 CTTACAGACTAAGCAGCAGCTGTAAGA 6979

RESULT 7

US-09-252-991A-863
; Sequence 863, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-16
; NUMBER OF SEQ ID NOS: 5
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 863
; LENGTH: 1788
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-863

Query Match 3.4%; Score 34; DB 4; Length 1788;
Best Local Similarity 56.1%; Pred. No. 0.41;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 688 ACAGACAGCGGTAAAGTTGTTCTACGGCTCAAGAACTCCGGCGGACGCTGTCGGTGG 747
DB 116 ACCGTACGCCGGCAAGCAGTCTTACTACTGCTTCGGCTGCGGCGCGGCGCAACGGG 175
QY 748 ATCGATATCGACATCGATGAGGACAAGCTTGAGTCTGCTCCAGGCGAGTTACGGAA 801
DB 176 CTCGGCTTCGTCATGGACCAAGCAGCTGAGTTCCTCCCGAGGCGGTCGAGGAA 229

RESULT 8

US-09-252-991A-885
; Sequence 885, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-16
; NUMBER OF SEQ ID NOS: 5
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 885
; LENGTH: 2097
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-885

Query Match 3.4%; Score 34; DB 4; Length 2097;
Best Local Similarity 56.1%; Pred. No. 0.46;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 688 ACAGACAGCGGTAAAGTTGTTCTACGGCTCAAGAACTCCGGCGGACGCTGTCGGTGG 747
DB 253 ACCGTACGCCGGCAAGCAGTCTTACTACTGCTTCGGCTGCGGCGCGGCAACGGG 312
QY 748 ATCGATATCGACATCGATGAGGACAAGCTTGAGTCTGCTCCAGGCGAGTTACGGAA 801
DB 313 CTCGGCTTCGTCATGGACCAAGCAGCTGAGTTCCTCCCGAGGCGGTCGAGGAA 366

RESULT 9

US-08-257-999-1/C
; Sequence 1, Application US/08257999
; Patent No. 5530195
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Currier, Thomas C.
; TITLE OF INVENTION: No. 5530195el Bacillus Thuringiensis Gene
; TITLE OF INVENTION: Encoding a Toxin Active Against Insects
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,999
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4003 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

```
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Bacillus thuringiensis
/ STRAIN: Kurstaki
/ INDIVIDUAL ISOLATE: CGB316
/ FEATURE:
/ NAME/KEY: promoter
/ LOCATION: 96..124
/ OTHER INFORMATION: /function= "putative promoter"
/ OTHER INFORMATION: region"
/ FEATURE:
/ NAME/KEY: RBS
/ LOCATION: 185..190
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 196..3723
/ OTHER INFORMATION: /product= "Full-length CryIE(c)"
/ OTHER INFORMATION: protein"
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 1191..1590
/ OTHER INFORMATION: /note= "This region of the CryIE(c)"
/ OTHER INFORMATION: DNA sequence encodes the amino acid sequence
/ OTHER INFORMATION: designated Sub-Sequence A"
/ Patent No. 5530195
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: 1591..2061
/ OTHER INFORMATION: /note= "This region of the CryIE(c)"
/ OTHER INFORMATION: DNA sequence encodes the amino acid sequence
/ OTHER INFORMATION: designated Sub-Sequence B"
/ Patent No. 5530195
/ US-08-257-999-1

Query Match 3.4%; Score 33.4; DB 1; Length 4003;
Best Local Similarity 54.5%; Pred. No. 1.1;
Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 862 CTGTTGAACCCCTACCGGAATACGAAATTATCTCCCGTCCACCGCCTATGGTCGAC 921
Db 2095 CTGTTTAAACCGATTGATTAGAGAGTAAACAGCGCATTCACCGCCCTTTTGCTC 2036
QY 922 GCTACTGCGTATGCTGCTCGGGAAGGTTTCCAGCGCATCAATTTTACGCA 981
Db 2035 TTCTAGATCATATTCCTCCTCAAGGTTACTTCTGCCGGAACAAATTCATTCGATCTA 1976
QY 982 TTT 984
Db 1975 TAT 1973

RESULT 10
US-08-319-387-1
; Sequence 1, Application US/08319387
; Patent No. 5543317
; GENERAL INFORMATION:
; APPLICANT: Shields, Malcolm S.
; APPLICANT: Francesconi, Stephen C.
; TITLE OF INVENTION: Microbial Degradation of Trichloroethylene,
; TITLE OF INVENTION: Dichloroethylenes and Aromatic Pollutants
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/319,387
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION NUMBER: US 08/167,457
/ FILING DATE: 15-DEC-1993
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/694,718
/ FILING DATE: 02-MAY-1991
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Whitlock, Ted W.
/ REGISTRATION NUMBER: 36,965
/ REFERENCE/DOCKET NUMBER: UWF-2
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 904-375-8100
/ TELEFAX: 904-372-5800
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 9785 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ US-08-319-387-1

Query Match 3.3%; Score 32.8; DB 1; Length 9785;
Best Local Similarity 48.0%; Pred. No. 3.2;
Matches 94; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 487 GAGTTCATATTAAGAGGGTACCCACAGCAGATTAGTCTTGCTTTTCGAAAATAGA 546
Db 6680 GAGCTCAATATTCGCGGGTTCGGGAGACAGGCGGTACATCATGACAACTG 6739
QY 547 AAAGAAAGCGCTCGTCTATTTTTCAGCGGACAAATGGGCACATCTTTCTCCGTCCAGG 606
Db 6740 GCGGTGGGGAACATCTCTGACGTGACGGGTCCCTATGCGCGGTTTTTCGTGGCAGGTCA 6799
QY 607 ACCGCCGGAAGAGTCTTTTGATTGGCGCGGTGCGGGCTCTCGTATGCGGCCCTATT 666
Db 6800 GCGGACCAAGCCATGGTTTTTATGSCAGTGTTTCAGSCCTGTGAGCCCTCGGTGATG 6859
QY 667 GCACGCGCTCGATGC 682
Db 6860 ATTCTCGACTTGTTC 6875

RESULT 11
US-09-621-976-18033
; Sequence 18033, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18033
; LENGTH: 474
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 16
; OTHER INFORMATION: n=a, g, c or t
; US-09-621-976-18033
```


Query Match 3.3%; Score 32.6; DB 4; Length 474;
Best Local Similarity 12.0%; Pred. No. 0.54; Mismatches 103; Indels 0; Gaps 0;
Matches 29; Conservative 109;

QY 276 CAAGTAAGTCCCAAGATCCCAATCTCAAGATGGAAGCGAGTGTGAGGTCCGGCG 335
DB 132 MWKRWYCCSSCMYTKGGGSMWTTTMMERRKKS YKRWTKGKKKKTMMMAAUCYTTWR 191
QY 336 GCTAACTCATGACCTGCTCGTGGGATTAGCGACTGATGGCCAGCAATTTCTCC 395
DB 192 SYWMMRRRAAAKTYTCMBSKTCCWACCCMCMRRARSCCWRSCMYTYMMCY 251
QY 396 CGGCCAGTTCTCGTCTAGTAGAGGAGAGCAGTTCGACGCGTGGTTCGCGATATCAAT 455
DB 252 YNMYKGRMYWNRGGMKRYMYMYKKSMWKGSCWKGAWARKTYYTWAUYTYTK 311
QY 456 GCGCAATTTAAAGACCCCGAGGAGGATGCGAGTTCATATTAGAGGGTACCCACAGG 515
DB 312 RMCCYMRKTYTCMMSYMRWRSWMTARGAWMCMYMYMYMAARKKYMMAAAREGG 371
QY 516 A 516
DB 372 W 372

RESULT 12
US-09-489-039A-6774/c
; Sequence 6774, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 6774
; LENGTH: 3972
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-6774

Query Match 3.3%; Score 32.2; DB 4; Length 3972;
Best Local Similarity 59.1%; Pred. No. 2.9; Mismatches 38; Indels 0; Gaps 0;
Matches 55; Conservative 0;

QY 181 GCACCTTTCTTCGCGAGATCGTGAGAGGCAACCGCCATCTTCGATCCAGTGGTTCGG 240
DB 1155 GTACTTAATGCCCGGACGCTGACGAGGTTCCGCCAGTTGGTGGCCAGCAGATCGG 1096
QY 241 CTCACAGCTCGGATCAAGTCCGAGTGCAG 273
DB 1095 CCGCCGCTGTGCGGCTGGAAGACGCGGTCTG 1063

RESULT 13
US-08-676-818-1
; Sequence 1, Application US/08676818
; Patent No. 6057136
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley Grant
; APPLICANT: Perkins, John B.
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Pero, Janice G.
; TITLE OF INVENTION: BIOTIN BIOSYNTHESIS IN BACILLUS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,818
FILING DATE: 08-JUL-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/239,430
FILING DATE: May 6, 1994
APPLICATION NUMBER: 08/084,709
FILING DATE: June 25, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 04599/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8478
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-676-818-1

Query Match 3.3%; Score 32.2; DB 3; Length 8478;
Best Local Similarity 46.9%; Pred. No. 4.8; Mismatches 100; Conservative 0; Indels 0; Gaps 0;

QY 681 GCGGAAACAGACAGCGGTAAAGTTGTTCTACGGCTCAAGAACTCCGGCGGACGCTGT 740
DB 2008 GCACGAGATGTCAGCCTGATCTGATGCTGCGGTAAAGGCAATTACAGGAGGCTATTT 2067
QY 741 TCGGTGGATCGATATGCATCATGAGGACAGCTTGAGTGTGTCAGGACGCTTACGGA 800
DB 2068 GCCAATTCGCGTTACGTTTGCCACTGAAGACATCTATAGGCATTCTATGATGATTATGA 2127
QY 801 AGACACGATAGCCTTTGGCAAGGCGCCACTGGTTTTATTTCATCAGGTTGTCCAGCAGC 860
DB 2128 AAACCTAAAAACCTTTTCCATGCCATTCCCTATACAGGCAATCAGCTTGGCTGCGGT 2187
QY 861 GCTGCTTGAACCTACCGGAATACGAAATTA 893
DB 2188 TGGCTTGAATAATCGSCATTATTGAATCTGA 2220

RESULT 14
US-09-407-549-1
; Sequence 1, Application US/09407549
; Patent No. 6303377
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley Grant
; APPLICANT: Perkins, John B.
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Pero, Janice G.
; TITLE OF INVENTION: BIOTIN BIOSYNTHESIS IN BACILLUS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts

```

; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/407,549
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/239,430
; FILING DATE: May 6, 1994
; APPLICATION NUMBER: 08/084,709
; FILING DATE: June 25, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/POCKET NUMBER: 04599/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8478
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-09-407-549-1
;
; Query Match 3.3%; Score 32.2; DB 4; Length 8478;
; Best Local Similarity 46.9%; Pred. No. 4.8;
; Matches 100; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
;
; QY 691 GCGGAAACAGACAGCGGTAAGTTGTTCTACGGCTCAAGAACTCCGCGCGACGGCTGT 740
; DB 2008 GCACGAGAATGCCAGCGCTGATCTGATGGCTGCGCGTAAAGGCATTACAGGAGGCTATTT 2067
;
; QY 741 TCGGTGGATCGATATCGACATCGATGAGGCAAGCTTGAGTCTGCCAGGCGATTACGGA 800
; DB 2068 GCCAATGCCGTTACGTTGGCCTGACGAGACATCTATAAGSCATTTATGATGATTAGA 2127
;
; QY 801 AGACAGGATAGCCTTTGGCAAGGCGCCACCTGGTTTTTATTCATCAGGTTGTCGACGAGC 860
; DB 2128 AAACCTAAACACCTTTTCCATGGCCATTCCTATACAGGCAATCAGCTTGCTGTGGGT 2187
;
; QY 861 GCTGCTTGAACCCCTACCGGAATACGAAATTA 893
; DB 2188 TGCCTTGAAATCTGGCATTATTGGAATCTGA 2220
;
; RESULT 15
; US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: PRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis

```

```

; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
; US-09-103-840A-2
;
; Query Match 3.3%; Score 32.2; DB 3; Length 4403765;
; Best Local Similarity 47.3%; Pred. No. 1.1e+02;
; Matches 97; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
;
; QY 746 GGATCGATATCGACATCGATGAGGACAAAGCTTGAGGTGCTCCAGGCACTTACGGGAAGACA 805
; DB 1825965 GAATCGTTGTCGACGGCGATGTTTGGCGCGTGAAGTGGTCCAGCGCGGCACCGAGGGGC 1826024
;
; QY 806 CGGATAGCCTTTGGCAAGGCCCCACTGTTTATTCATCAGTTTCTCGACGAGCGCTGC 865
; DB 1826025 TGGCCTCGCTGGTCCAGCGCTTGGTTCGGACATCTCTGTCACAGGAGCGCTGGACC 1826084
;
; QY 866 TTGAAACCCCTACCGGAATACGAAATTTATCTTCCGCGTCCACCGCTATGTTGTCGACGCTA 925
; DB 1826085 GGCAGGCGTTGGCGCGCCCAAGGCGTTTCGAGATGACGAGTCCGCGCGTGTGCTCAACGGAA 1826144
;
; QY 926 CTGTCGCTATGCTGCTCGGCAAGGG 950
; DB 1826145 TCGTCACCGCGCTGGTCCGCGCG 1826169
;
; Search completed: June 27, 2004, 10:02:59
; Job time : 96.4159 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 07:56:42 ; Search time 456.449 Seconds

(without alignments)

9936.011 Million cell updates/sec

Title: US-10-658-691-3

Perfect score: 990

Sequence: 1 atgtaccactcaaaattga.....atttgagcatttttctaa 990

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3017426 seqs, 2290544650 residues

Total number of hits satisfying chosen parameters: 6034852

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
- 13: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq*
- 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq*
- 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq*
- 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
- 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	990	100.0	990	17	US-10-658-691-3
2	54.6	5.5	12808	10	US-09-843-250-10
3	53	5.4	9706	10	US-09-843-250-5
4	49.8	5.0	4355	10	US-09-843-250-7
5	45.6	4.6	14462	10	US-09-843-250-9
6	37	3.7	1029	13	US-10-282-122A-13049
7	37	3.7	3673778	15	US-10-312-841-1
8	36.4	3.7	2394	13	US-10-282-122A-33544
9	35.8	3.6	5331	15	US-10-223-371B-1
10	35.6	3.6	632	9	US-09-864-761-20742
11	35.6	3.6	945	13	US-10-282-122A-12871
12	35.6	3.6	3981	9	US-09-864-761-3981
13	35.4	3.6	1240	16	US-10-260-238-3074
14	34.8	3.5	65042	15	US-10-229-124-3

15	34.6	3.5	1038	15	US-10-213-878-11	Sequence 11, Appl
16	34.6	3.5	1038	15	US-10-214-059-11	Sequence 11, Appl
17	34.6	3.5	1041	15	US-10-213-878-21	Sequence 21, Appl
18	34.6	3.5	1041	15	US-10-214-059-21	Sequence 21, Appl
19	34.6	3.5	12591	15	US-10-213-878-6	Sequence 6, Appl
20	34.6	3.5	12591	15	US-10-214-059-6	Sequence 6, Appl
C 21	34	3.4	888	13	US-10-398-916-25	Sequence 25, Appl
C 22	34	3.4	900	13	US-10-398-916-12	Sequence 12, Appl
C 23	34	3.4	1995	13	US-10-282-122A-30013	Sequence 30013, A
C 24	33.8	3.4	586	13	US-10-027-632-129703	Sequence 129703,
C 25	33.8	3.4	586	16	US-10-027-632-129703	Sequence 129703,
C 26	33.8	3.4	1185	16	US-10-369-493-41550	Sequence 41550, A
C 27	33.8	3.4	1711	16	US-10-369-493-41550	Sequence 41550, A
C 28	33.2	3.4	1023	13	US-10-389-647-172	Sequence 172, App
C 29	33	3.3	750	15	US-10-184-644-104	Sequence 104, App
C 30	33	3.3	750	15	US-10-184-634-104	Sequence 104, App
C 31	33	3.3	912	13	US-10-282-122A-15013	Sequence 15013, A
C 32	33	3.3	2117	13	US-10-425-114-13860	Sequence 13860, A
C 33	32.8	3.3	888	16	US-10-369-493-33274	Sequence 33274, A
C 34	32.6	3.3	696	13	US-10-142-426-354	Sequence 354, App
C 35	32.6	3.3	696	15	US-10-123-155-354	Sequence 354, App
C 36	32.6	3.3	696	15	US-10-146-731-354	Sequence 354, App
C 37	32.6	3.3	696	15	US-10-140-472-354	Sequence 354, App
C 38	32.6	3.3	696	15	US-10-141-761-354	Sequence 354, App
C 39	32.6	3.3	696	15	US-10-142-885-354	Sequence 354, App
C 40	32.6	3.3	696	15	US-10-158-790-354	Sequence 354, App
C 41	32.6	3.3	696	16	US-10-137-871-354	Sequence 354, App
C 42	32.6	3.3	696	16	US-10-140-923-354	Sequence 354, App
C 43	32.6	3.3	696	16	US-10-141-756-354	Sequence 354, App
C 44	32.6	3.3	696	16	US-10-141-759-354	Sequence 354, App
C 45	32.6	3.3	696	16	US-10-140-805-354	Sequence 354, App

ALIGNMENTS

RESULT 1

US-10-658-691-3
; Sequence 3, Application US/10658691
; Publication No. US20040110258A1
; GENERAL INFORMATION:
; APPLICANT: Kayser, Kevin J.
; APPLICANT: Kilbane, John J.
; TITLE OF INVENTION: Method for Metabolizing Carbazole in Petroleum
; FILE REFERENCE: GTI-1512
; CURRENT APPLICATION NUMBER: US/10/658,691
; CURRENT FILING DATE: 2003-09-09
; PRIOR APPLICATION NUMBER: US 60/409,562
; PRIOR FILING DATE: 2002-09-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 990
; TYPE: DNA
; ORGANISM: Pseudomonas resinovorans
; US-10-658-691-3

Query March 100.0%; Score 990; DB 17; Length 990;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 990; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGTACCAACTCAAAATTAAGGGCAGGACCTGGGCTCAGGAGAGCCCTG	60
Db	1	ATGTACCAACTCAAAATTAAGGGCAGGACCTGGGCTCAGGAGAGCCCTG	60
Qy	61	TTGGTCTCAGCACTTCTAATGATCGGATTTCCGTACAGTGTGTCATCGGAGGTTC	120
Db	61	TTGGTCTCAGCACTTCTAATGATCGGATTTCCGTACAGTGTGTCATCGGAGGTTC	120
Qy	121	GGAGTATGCAAAATTCAGTACTTCGAAGGGAATGTCCAATCAATGTGGCCGATGTCCA	180
Db	121	GGAGTATGCAAAATTCAGTACTTCGAAGGGAATGTCCAATCAATGTGGCCGATGTCCA	180

181 GGACTTTCTTCGCGAGATCGTGAGAGGCGCAACCCCATCTTGTGATGCCAGTGGTTCG 240
Db 181 GGACTTTCTTCGCGAGATCGTGAGAGGCGCAACCCCATCTTGTGATGCCAGTGGTTCG 240
Qy 241 CTCTAGACCTCGGATCAAAAGTCGAGTGCAGGCAAGATGCTCCCAACGATTCGAATC 300
Db 241 CTCTAGACCTCGGATCAAAAGTCGAGTGCAGGCAAGATGCTCCCAACGATTCGAATC 300
Qy 301 TCAAGAAATCGAAGCGGAAGTCTGTGAGTCCGGCGCTAACTCATGACTGCTGTCGCTG 360
Db 301 TCAAGAAATCGAAGCGGAAGTCTGTGAGTCCGGCGCTAACTCATGACTGCTGTCGCTG 360
Qy 361 CGATTACGCACCTGATGGGCGCAGCAAAATTTCTCCCGCGCAGTCTGCTAGTAGAGCA 420
Db 361 CGATTACGCACCTGATGGGCGCAGCAAAATTTCTCCCGCGCAGTCTGCTAGTAGAGCA 420
Qy 421 GAGCAGTTCCCGAGGCGTGTTCGCGCATATTCATGAGGGAATTTAAAGAACCCCGAAGGC 480
Db 421 GAGCAGTTCCCGAGGCGTGTTCGCGCATATTCATGAGGGAATTTAAAGAACCCCGAAGGC 480
Qy 481 ATATGAGAGTCTATATTAAGAGGTPACCAAGAGGATTTAGTCTTGGCTTTTCGAA 540
Db 481 ATATGAGAGTCTATATTAAGAGGTPACCAAGAGGATTTAGTCTTGGCTTTTCGAA 540
Qy 541 AATAGAAAGAGGCGCTGCTATTTTTCGCGGACCAATGGGACATCTTCTTCGCT 600
Db 541 AATAGAAAGAGGCGCTGCTATTTTTCGCGGACCAATGGGACATCTTCTTCGCT 600
Qy 601 CCAGGACCGCGCGAAGAGTCTTTGCAATGGCGCGTGCCTGATCGGCGC 660
Db 601 CCAGGACCGCGCGAAGAGTCTTTGCAATGGCGCGTGCCTGATCGGCGC 660
Qy 661 GCTATTGCAAGGCGCTGATGGCGAAGACAGCAAGCGGTAAGTGTCTACGGCTCA 720
Db 661 GCTATTGCAAGGCGCTGATGGCGAAGACAGCAAGCGGTAAGTGTCTACGGCTCA 720
Qy 721 AGAACTCCGCGCGACGCTGTTCGTTGATGATCGATCGATCGATCGATCGATCGATCG 780
Db 721 AGAACTCCGCGCGACGCTGTTCGTTGATGATCGATCGATCGATCGATCGATCGATCG 780
Qy 781 GTCTCCAGGAGTTCAGGAGACAGGATAGCTTTGCAAGGCGCCTGCTTTTATT 840
Db 781 GTCTCCAGGAGTTCAGGAGACAGGATAGCTTTGCAAGGCGCCTGCTTTTATT 840
Qy 841 CATCAGGTTGTCGACGACGCTGCTTTGAAACCTACCGAATACGAATTTATCTTGC 900
Db 841 CATCAGGTTGTCGACGACGCTGCTTTGAAACCTACCGAATACGAATTTATCTTGC 900
Qy 901 GTTCCACCGCTATGTCGACGCTACTGTCGATGCTGCTCGGCAAGGTTTCCAGC 960
Db 901 GTTCCACCGCTATGTCGACGCTACTGTCGATGCTGCTCGGCAAGGTTTCCAGC 960
Qy 961 GATCAATTCATTTGACGCAATTTTCTAA 990
Db 961 GATCAATTCATTTGACGCAATTTTCTAA 990

RESULT 2

US-09-843-250-10
; Sequence 10, Application US/09843250
; Publication No. US20030022335A1
; GENERAL INFORMATION:
; APPLICANT: Parales, R.
; APPLICANT: Gibson, D.
; APPLICANT: Resnick, S.
; APPLICANT: Lee, K.
; TITLE OF INVENTION: No. US20030022335A1 naphthalene dioxygenase and methods for the
; FILE REFERENCE: 875.006US2
; CURRENT APPLICATION NUMBER: US/09/843,250
; CURRENT FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: PCT/US99/25079
; PRIOR FILING DATE: 1999-10-26
; PRIOR APPLICATION NUMBER: US 60/105,575

PRIOR FILING DATE: 1998-10-26

NUMBER OF SEQ ID NOS: 65

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10

LENGTH: 12808

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: A modified DNA molecule encoding valine at the

OTHER INFORMATION: Position corresponding to the F352 amino acid in

OTHER INFORMATION: NDO.

US-09-843-250-10

Query Match 5.5%; Score 54.6; DB 10; Length 12808;
Best Local Similarity 50.4%; Pred. No. 2e-07;
Matches 191; Conservative 0; Mismatches 179; Indels 9; Gaps 2;

Qy 303 AAGNATGGAGCGGAAGTGTTCAGGTCGCGGCGCTAACTCATGACTGCTGTCGCTGCG 362
Db 1101 AATCATCAAGGATGTTGTGCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCG 1160
Qy 363 ATTACGCACCTGATGGGCGCAGCAAAATTTCTCCCGCGCAGTTCGCTAGTAGAGGCA 422
Db 1161 CATTCGCTCGCCAAAGCCCTTCGAGTTCACCCGCGACAGTAC-----GCGATGCTACA 1214
Qy 423 GCAGTTCGCGGCGTGTTCGCGCATATTCATGAGGCAATTTAAAGAACCCCGAAGCAT 482
Db 1215 GTTCAGTCCCGAAGCATGTGCGTCCATATTCATGAGTTCGCGGCTGCGGCTGCGGCTGCG 1274
Qy 483 ATGGGAGTTCATATTAAGAGGCTAGCCACAGGACGATTTAGTCTTGGCTTTTCGAAA 542
Db 1275 ---GGATTCATATCCGCAAGTGCAGGCGCGGCTGTCACGAGTATATTTTCGAGCA 1331
Qy 543 TAGAAAAGAGGCGCTGCTATTTTTCGCGGACCAATGGGCAATCTTCTTCGCTCC 602
Db 1332 CGTCCGGAAGTACAGCATTAAGTGTGAGTGGGCTCTTGTGACGCGCTATCTGCGTCA 1391
Qy 603 AGGACCGCGCGAAGAGTCTTTGCAATGGCGCGGTCGCGGCTCTGCTATGCGGCGC 662
Db 1392 GGCTCACACCGCGCGGATGCTGTGTGGCGCGGCGGAGTCCGACCGGTGCTGTC 1451
Qy 663 TATTGACGCGCTCGATG 681
Db 1452 GATTGTTCCGCGCGCTG 1470

RESULT 3

US-09-843-250-5

; Sequence 5, Application US/09843250

; Publication No. US20030022335A1

; GENERAL INFORMATION:

; APPLICANT: Parales, R.

; APPLICANT: Gibson, D.

; APPLICANT: Resnick, S.

; APPLICANT: Lee, K.

; TITLE OF INVENTION: No. US20030022335A1 naphthalene dioxygenase and methods for the

; FILE REFERENCE: 875.006US2

; CURRENT APPLICATION NUMBER: US/09/843,250

; CURRENT FILING DATE: 2001-04-26

; PRIOR APPLICATION NUMBER: PCT/US99/25079

; PRIOR FILING DATE: 1999-10-26

; PRIOR APPLICATION NUMBER: US 60/105,575

; PRIOR FILING DATE: 1998-10-26

; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5

LENGTH: 9706

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: A modified DNA molecule encoding valine at the

OTHER INFORMATION: Position corresponding to the F352 amino acid in

OTHER INFORMATION: NDO.

US-09-843-250-5
Query Match 5.4%; Score 53; DB 10; Length 9706;
Best Local Similarity 50.1%; Pred. No. 6.4e-07;
Matches 190; Conservative 0; Mismatches 180; Indels 9; Gaps 2;

QY 303 AAGATGGAAGCGGAGTTGTTGAGGTCCGGGCGCTAACTCATGACCTCTGTCCGTGCG 362
DB 1193 AATCATCAAGGGCACAGTGTGCGAGTCCGACCTCAGTATCCGTCGCTTACG 1252

QY 363 ATTACGCACTGATGGCCACCAATTTCTCCCGGCCAGTTCTGCTAGTAGGCGCA 422
DB 1253 CGTACGCTCTCAAGCCCTTCGAGTTCTCACC CGGACAGTACGCGACT-----GCA 1306

QY 423 GCAGTTTCCAGGGGTGTTCCGCATATTCATGCGGGAATTTAAAGAACCCCGAAGGCAT 482
DB 1307 GTTCAGCCCTGAGCATGCGGCTCGTATTCAATGCGCAGTTTCCAGATGACCAAGAAAT 1366

QY 483 ATGGGAGTTCTATTAAGAGGTACCCACAGACGATTTAGTCCCTTGGCTTTTCGAAA 542
DB 1367 ---GGAGTTCCACATACGCAAGGTGCGGGTGGCGCGTACAGAGTATGTTTTTCGAACA 1423

QY 543 TAGAAGAAAGGGCGCTCGTCTATTTTTCAGCGGACCAATGGGCACATCTTTCTTCGTCG 602
DB 1424 CGTCCGGAAGGTACAGCATCAAGTTGACGGGCTCTTGTGACGGCTTATCTAGCTCA 1483

QY 603 AGGACCGGCGGAAAGAGTCTTTGCATTTGCGCGGCTGCGGCGCTCTGATGCGGCCGCG 662
DB 1484 GAAGCACACCGGACCGATGCTGTGTGATGTTGGCGGACCGGACTCGCACCGTGTCTGTC 1543

QY 663 TATTGACGCGCTCGCATG 681
DB 1544 GATTGTCGGGCGGCTG 1562

RESULT 4
US-09-843-250-7
; Sequence 7, Application US/09843250
; Publication No. US20030022335A1
; GENERAL INFORMATION:
; APPLICANT: Parales, R.
; APPLICANT: Gibson, D.
; APPLICANT: Resnick, S.
; APPLICANT: Lee, K.
; TITLE OF INVENTION: No. US20030022335A1 naphthalene dioxygenase and methods for the
; FILE REFERENCE: 875.006US2
; CURRENT APPLICATION NUMBER: US/09/843,250
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: PCT/US99/25079
; PRIOR FILING DATE: 1999-10-26
; PRIOR APPLICATION NUMBER: US 60/105,575
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 4355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A modified DNA molecule encoding valine at the
; OTHER INFORMATION: position corresponding to the F352 amino acid in
; OTHER INFORMATION: NDO.
US-09-843-250-7

Query Match 5.0%; Score 49.8; DB 10; Length 4355;
Best Local Similarity 50.1%; Pred. No. 5.8e-06;
Matches 182; Conservative 0; Mismatches 172; Indels 9; Gaps 2;

QY 319 GTTGTGAGGTCCGGGCGCTAACTCATGACCTGCTGCGGATTAACGCACTGATGG 378
DB 1159 GTGTGCGCGGTGAGTGTGCGCCCACTCAGATATCGCTACGCTACGCTCGCTAAG 1218

QY 379 CCAGCAATTTCTCCCGCGCCGCTTCTCTAGTAGAGGACAGACGATTGCCAGGCGTG 438

DB 1219 CCCTTCGAGTTCTCACCCGGACAGTACGCGACATT-----GCAGTTCACTCTGAGCAT 1272
QY 439 GTTCCGCGATATTCAATGGCGAAATTTAAAGAACCCCGAAGCATTATGGAGTTCTATATT 498
DB 1273 GCGGTCCTGATATTCAATGGCAGGTCTGCCAGATACCAAGAAAT---GGAGTTCCACATA 1329
QY 499 AAGAGGGTACCCACAGGACGATTTAGTCTTTGGCTTTTCGAAAAATAGAAAAAGAGGCGCT 558
DB 1330 CGCAAGGTGCGGGTGGCGGTAACGGAGTATGTTTTCGAGCAGCTCCCGGAAGGTACA 1389
QY 559 CGTCTATTTTTCAGCGGACCAATGGGCACATCTTTCTTCGTCAGGACCGGCGGAAG 618
DB 1390 AGCATCAAGTTGAGCGGCGCACTTTGGTACGGCTTATTTGCGTCAGAACACACACCGGGCG 1449
QY 619 AGTCTTTTCATTTGCGGCGGTGCGGGCTCTCTGATGCGGCGCTATTTGACCGGCGCTCG 678
DB 1450 ATGCTCTGTGTGGCGGTGGGACCGGACTAGCACCGGTGCTGCTGATTTGTCGGGCGCG 1509
QY 679 ATG 681
DB 1510 CTG 1512

RESULT 5
US-09-843-250-9
; Sequence 9, Application US/09843250
; Publication No. US20030022335A1
; GENERAL INFORMATION:
; APPLICANT: Parales, R.
; APPLICANT: Gibson, D.
; APPLICANT: Resnick, S.
; APPLICANT: Lee, K.
; TITLE OF INVENTION: No. US20030022335A1 naphthalene dioxygenase and methods for the
; FILE REFERENCE: 875.006US2
; CURRENT APPLICATION NUMBER: US/09/843,250
; PRIOR FILING DATE: 2001-04-26
; PRIOR APPLICATION NUMBER: PCT/US99/25079
; PRIOR FILING DATE: 1999-10-26
; PRIOR APPLICATION NUMBER: US 60/105,575
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 14462
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: A modified DNA molecule encoding valine at the
; OTHER INFORMATION: position corresponding to the F352 amino acid in
; OTHER INFORMATION: NDO.
US-09-843-250-9

Query Match 4.6%; Score 45.6; DB 10; Length 14462;
Best Local Similarity 52.0%; Pred. No. 0.00034;
Matches 102; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 486 GGAGTTCTATATTAAAGAGGTACCCACAGAGGATTTAGTCTCTGGCTTTTCGAAAAATAG 545
DB 4235 GGAGTTCCACATACGCAAGGTGCGGGTGGCGCGTAAGTATGTTTTCGACACGT 4294
QY 546 AAAAGAGGCGCTCTGCTATTTTTCAGCGGACCAATGGGCACATCTTTTTCGTCGAGG 605
DB 4295 CCGCAAGGTACCAAGCATCAAAATTGACGGGCGCACTTGGTACGGCTTATTTACGTCAAG 4354
QY 606 GACCGGCGGCAAGAGTCTTTTTCATTTGGCGGCGGTGCGGCGTCTCTGATGCGGCGCTAT 665
DB 4355 CCACACCGGCGCGATGCTCTGTGTGGCGGTGGACCGGACTAGCACCGGTCTGTGAT 4414
QY 666 TCACGCGCTCGATG 681
DB 4415 TATTGCGGCGGCTG 4430

RESULT 6
US-10-282-122A-13049
; Sequence 13049, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13049
; LENGTH: 1029
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-282-122A-13049

Query Match 3.7%; Score 37; DB 13; Length 1029;
Best Local Similarity 47.8%; Pred. No. 0.097;
Matches 140; Conservative 0; Mismatches 150; Indels 3; Gaps 1;
QY 433 GCGCTGTTCCGCGATATCAATGGCGAATTTAAAGAACCCGGAAGGCATATGGGAGTTC 492
Db 433 GCGAAGCGCGAGCTATTCATGGCGAATGCCGCCACACGGAAGCGCGATCGAATG 492
QY 493 TATATTAAGAGGTACCCAGACGATTTAGTCTTGCTTTTCGAAATAGAAAGAA 552
Db 493 CATATCCGTACATGCGCGGTGTGCTTTACCGATCATGTGTTCACACCATGAAAGAG 552
QY 553 GCGCTCGTCTATTTTTCGCGGACCAATGGGCACATCTTCTTCGTCGAGGACCGGC 612
Db 553 CGGACATCTGCGCTTCGAAGCGCGCTCGGCACGTTCTTCTGCGCAAGATTCGGAC 612
QY 613 CGAAGAGTCTTTTCATTTGCGGCGGTGCGGCGCTCT---CGTATCGGCGCGTATTCGA 669
Db 613 AAGCGATCTGCTGCTCGGCTCGGCTCGGCTCGGCGCTTGAAGCGATCATCGAG 672
QY 670 CGGCTCTGATCGCGAAGACAGACCGGTAAGTTGTTCTACGGCTCAAG 722

Db 673 CACGCGTGTTCAGAACCTGAACCGCGCATGACGCTGTACTGGGCGCAGC 725
RESULT 7
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MH
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

Query Match 3.7%; Score 37; DB 15; Length 3673778;
Best Local Similarity 57.3%; Pred. No. 7.1;
Matches 67; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 738 TGTTCCGTGGATCGATATCGACATCGATGAGGACACGCTTGAGTCGTCAGGAGTTAC 797
Db 878233 TGTTAGTTAAATCGATGATTTTGAAGATTATAAGCGGAGGTGTGGAGGTTTGGG 878292
QY 798 GGAAGACACGATAGCCCTTTGGCAAGGCCCCACTGGTTTTATTTCATCAGGTTGTGCA 854
Db 878293 TGAAGTTATGTTAGTCGTTTCGAGAGTGAAGTTTTTTTTTTAGAGTTTTTCGA 878349

RESULT 8
US-10-282-122A-33544
; Sequence 33544, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33544
; LENGTH: 2394
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
US-10-289-132A-33544

```

	Query Match	3.7%;	Score 36.4;	DB 13;	Length 2394;
	Best Local Similarity	50.0%;	Pred. No. 0.25;		
	Matches 91;	Conservative 0;	Mismatches 91;	Indels 0;	Gaps 0;
QY	310	GAACGCGAAGTGTGTTGAGGTCCGGCGCTAACTCATGACCTGCTGTCGTCGCATTACGC	369		
Db	364	GAATCGAAGTGTGTTGCGCAGCTCTGCTCTAATTCGACGAGTACGTGCATCTGGC	423		
QY	370	ACTGATGGCGCAGCAAAATTTCCCTCCCGGCCAGTTCTTCCTTAGTAGAGCGACAGTTG	429		
Db	424	AAGAAAGTCCCGCAGAACTCCTCTCTCGCTCAACAGCATGTGATGAGCCAGCGCGCTG	483		
QY	430	CCAGGCGTGGTTCGGGCATATTCAATGSCGAATTTAAAGAACCCCGAGGCGCATATGGAG	489		
Db	484	GTAGACACCATGGCGGCATATGGCGCTGAAATTCGAGCAGACGAGAAATCCTTGAG	543		
QY	490	TT 491			
Db	544	AT 545			

RESULT 9
 US-10-223-371B-1
 ; Sequence 1, Application US/10223371B
 ; Publication No. US20030170877A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yano, Tetsuya; No. US20030170877A1oto, Tsuyoshi; Imamura, Takeeshi; Canon
 ; APPLICANT: Kabushiki Kaisha
 ; TITLE OF INVENTION: DNA Fragment Carrying Toluene Monooxygenase Gene, Recombinant
 ; TITLE OF INVENTION: Plasmid, Transformed Microorganism, Method for Degrading Halogen
 ; TITLE OF INVENTION: Aliphatic Hydrocarbon Compounds and Aromatic Compounds, and Meth
 ; TITLE OF INVENTION: Environmental Remediation
 ; FILE REFERENCE: 03500.014074.1
 ; CURRENT APPLICATION NUMBER: US/10/223.371B
 ; CURRENT FILING DATE: 2003-03-13
 ; PRIOR APPLICATION NUMBER: US/03/453,956
 ; PRIOR FILING DATE: 1999-12-03
 ; NUMBER OF SEQ ID NOS: 14
 ; SEQ ID NO 1
 ; LENGTH: 5331
 ; TYPE: DNA
 ; ORGANISM: Ralstonia eutropha
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (200)..(445)
 ; OTHER INFORMATION: comK
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (446)..(1437)
 ; OTHER INFORMATION: comL
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1470)..(1739)
 ; OTHER INFORMATION: comW
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1777)..(3324)
 ; OTHER INFORMATION: comY

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3341)...(3697)
; OTHER INFORMATION: tomQ
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3737)...(4795)
; OTHER INFORMATION: comp
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4797)...(5132)
; OTHER INFORMATION: tomQ
US-10-223-371B-1

Query Match          3.6%; Score 35.8; DB 15; Length 5331;
Best Local Similarity 43.4%; Pred. No. 0.61;
Matches 279; Conservative

Qy      48 AGGGAAGAGCCTGTGGTCTCACACTTGCTAATGTTACGGATTTCGGTACGAGTGTGC 107
Db      3794 AGGGCAGACCATTCTCGATCGCGCGTGTGCCCAAGCATCTATATCCGCATGCCCTGTG 3853
Qy     108 ATCGGAGGTTGCGGAGTATGCAAATTCGAGTTACTCGAAGGGAATGTCCAATCAATGTG 167
Db    3854 CCACGGCGCTGTGCGGACCTGCAAAGTCTCGTCTCTCGACGGCAGGCCGACCTCGGCGCA 3913
Qy     168 GCCGGATGTCACAGGACTTTCTTCGCGAGATCGTGAGAAGGGCAACCGCATCTTGCATG 227
Db    3914 GGCACACCGCTTCGGTTGATGTAATTCGAGCCCGAGAGGGCAGCGCTGCGGTGCTG 3973
Qy     228 CCAGTGGTTGCGCTCTCAGACCTCGCGATCAAAGTCGCAGTCGAGACAAGTACGTCCC 287
Db    3974 CGCGAGCTCGAGGCCGATACCAACCATCGAGGCCGATGTCACAAGACCCGACGCGCA 4033
Qy     288 AACGATTCGAATCTCAAGATGGAAGCGGAATGTTGAGTTCGGGCGCTAACTACTATGA 347
Db    4034 GATATCCCGTTCGGGATTTTCAGGCCGACGTGATGTGATCGAACGCTCACCCGAC 4093
Qy     348 CCTGCTCTCGTTCGATTAGCACTGATGGCCAGCAAAATTTCTCCCCCGCCAGTTCTG 407
Db    4094 CATCAAGCGATCCGCTCGSGCTCGCGAGCCGATGCGTTTTCCAGCGCGCCAGTACGT 4153
Qy     408 CCTAGTAGGCGAGAGCAGTGTCCAGCGCTGTTTCGGCATATTCATGCGGAATTTAAA 467
Db    4154 CCAGTTGAG---ATCCCGGCTGGGCCAGACCCCGCTTTCTCGANTGCCAACCGGCC 4210
Qy     468 GAACCCCGA-----AGGCATATGGGAGTTCTATATTAAAGAGGTACCCACAGGACG 518
Db    4211 GGCGGAGCTCGCGCGACCGCGAGATCGAGCTGAACGTGCGCGAGGTGCGGGCGGCGCT 4270
Qy     519 ATTATGTCCTTGGCTTTTCGAAATAAGAAAGGCGCTCGTCTATTTTTHGCGGAC 578
Db    4271 TGGCACCAGCTACCTGCACGAGCAGTTCGCGCGCGGGGATCGCGTGTGCTTGTCCGAC 4330
Qy     579 AATGGGCACATCTTCTTCGCTCCAGGACCGGCCGAAAGAGTCTTTGCAATTCGCGCGG 638
Db    4331 CTATGGCGGCTCTTTGTTGCGCGCTGCGCGCGGCTTCCGATGATCTTATGGGGGCGG 4390
Qy     639 TGCGGGCTCTCGTATGCGGCCGCTATTTCACGCGGCTCGATG 681
Db    4391 CTGCGGGCTCTCGAGCCCGGCTCCATGATCTCGACCTGCTG 4433

```

```

RESULT 10
US-09-864-761-20742/c
; Sequence 20742, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OR INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

```


;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;; FILE REFERENCE: Aeomica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 20742
;; LENGTH: 632
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AE000658.1
;; OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 8.5
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7.1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.3
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.8
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 11
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.3
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 11
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.6
;; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 15
;; OTHER INFORMATION: NT HIT: G55032026, EVALUE 0.00e+00
;; OTHER INFORMATION: SWISSPROT HIT: Q09028, EVALUE 1.00e-113
;; OTHER INFORMATION: EST_HUMAN HIT: BE537587.1, EVALUE 0.00e+00
US-09-864-761-20742

Query Match 3.6%; Score 35.6; DB 9; Length 632;
Best Local Similarity 52.7%; Pred. No. 0.24; Indels 0; Gaps 0;
Matches 77; Conservative 0; Mismatches 69;

Qy 692 ACAAGCGGPAAGTTGTTTACGGCTCAAGAACTCCGGCGACGCTGTTCGGTGGATCG 751

Db 254 AGATTCCAGAAAACATCTTCTACTGCTGTATGCCCGTAAGATGCTCTTGCATCC 195

Qy 752 ATATCGACATCGATAGGCAAGCTTGAGTGTCTCCAGCAGTTACGGAGACCGGTA 811

Db 194 ACTTTCCCTCTTTGGAATGGCACTGATGTCCACAGGACAGACAGTGTGTCATCTGAA 135
Qy 812 GCCTTTGGCAAGGCCCCACACTGGTTTT 837
Db 134 GCCTAAGTAAATGCCACACTGAGATT 109

RESULT 11

US-10-282-122A-12871
; Sequence 12871, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

;; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;; FILE REFERENCE: ELITRA.034A
;; CURRENT APPLICATION NUMBER: US/10/282,122A
;; CURRENT FILING DATE: 2003-02-20
;; PRIOR APPLICATION NUMBER: 60/191,078
;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848
;; PRIOR FILING DATE: 2000-05-23
;; PRIOR APPLICATION NUMBER: 60/207,727
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/230,335
;; PRIOR FILING DATE: 2000-09-06
;; PRIOR APPLICATION NUMBER: 60/230,347
;; PRIOR FILING DATE: 2000-09-09
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/267,636
;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 12871
;; TYPE: DNA
;; LENGTH: 945
;; ORGANISM: Burkholderia fungorum
US-10-282-122A-12871

Query Match 3.6%; Score 35.6; DB 13; Length 945;
Best Local Similarity 50.6%; Pred. No. 0.29; Indels 0; Gaps 0;
Matches 86; Conservative 0; Mismatches 84;

Qy 131 AATTCGAGTTACTCGAAGGGAATGTCCAATCAATGTGCCCGGATGCTCCAGGACTTTCTT 190

Db 302 AACTCGCGCTCGAAGCAGCGGTGCTCTCCAGTTTCGGAAGCGGGCGGCTGCGTA 361

Qy 191 CGCAGATCGTGAGAGAGGCAACCGCCATCTTGCATGCCAGTGGTTCGCTCTCAGACC 250

Db 362 CCGCGGAATTTCTCGCGTGGCCCGGCAATTTTCGGGAAAGCGGCGTCCCGGAAAGTGC 421

Qy 251 TGCGGATCAAAAGTCGCGAGTGCAGGACAAAGTACGTCCCAAGATTCCAATC 300

Db 422 TGGCGAGCGCGTCTCGCGTGGGCAAGGAGCCATTACGACCTGATC 471

Db 197 GCCCGTCGGCCCTTGCACCTTGAGGGCGACGAACCTGCGCAGCGCTTTTCGC---CTGCGCT 253
Qy 227 GCCAGTGCCTTGGCGCTCTCAGACCTCGGATCAAAAGTCGACGTGC 271
Db 254 GCCAGGCCATTCCGCGCTCGGATCTGACAATCGCGGTTGATGCGC 298

Search completed: June 27, 2004, 13:08:52
Job time : 467.449 secs

RESULT 14
US-10-229-124-3
; Sequence 3, Application US/10229124
; Publication No. US20030013168A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THERIOF
; FILE REFERENCE: CL001139DIV
; CURRENT APPLICATION NUMBER: US/10/229,124
; CURRENT FILING DATE: 2002-08-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 65042
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(65042)
; OTHER INFORMATION: n = A,T,C or G
US-10-229-124-3

Query Match 3.5%; Score 34.8; DB 15; Length 65042;
Best Local Similarity 62.8%; Pred. No. 5.2;
Matches 54; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
Qy 534 TTTCGAAATAGAAAAGAGCGCTCGTCTATTTTTCGCGGACCAATGGGACATCTTT 593
Db 6894 TTTTGAATAAGAAAAGAGAGAGTCAACAGTAGAGACTAGACCTCTGGCAATACAAT 6953
Qy 594 CTTCCGTCACGAGCGCGCGGAAGA 619
Db 6954 CTTGAGACTAAGACCAGCTGTAAAGA 6979

RESULT 15
US-10-213-878-11
; Sequence 11, Application US/10213878
; Publication No. US20030073206A1
; GENERAL INFORMATION:
; APPLICANT: Bramucci, Michael
; APPLICANT: Nagarajan, Vasantha
; APPLICANT: Thomas, Stuart
; TITLE OF INVENTION: Use of Xylene Monooxygenase for the Oxidation of Substituted
; TITLE OF INVENTION: Monocyclic Aromatic Compounds
; FILE REFERENCE: CL1662 US NA
; CURRENT APPLICATION NUMBER: US/10/213,878
; CURRENT FILING DATE: 2002-08-07
; PRIOR APPLICATION NUMBER: 60/311,490
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 11
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Sphingomonas sp.
US-10-213-878-11

Query Match 3.5%; Score 34.6; DB 15; Length 1038;
Best Local Similarity 50.2%; Pred. No. 0.69;
Matches 113; Conservative 0; Mismatches 109; Indels 3; Gaps 1;
Qy 47 CAGGGAAGAGCCTTTGGTCTCAGCACTTGTCTAATGCTATCGGATTTCCGTACGAGTGTG 106
Db 77 CGGGCAAGACACTTCTTGAAGCGATGCTGGACCGCGGTCTGGCCATGCGCGCAGATTGCA 136
Qy 107 CATCGGAGGTTGGGAGTATGCAAAATTCGAGTTACTCGAAGGGAATGTCCAATCAATGT 166
Db 137 AGTTGGCTCGTGGGTACCTGCAAGTTCAAGCTCGTGTGCAAGATCGGCAATTGA 196
Qy 167 GGCCGGATGCTCCAGGACTTTCTTCGGGAGATCTGTGAAGAGGGCAACCGCCATCTTGCAT 226

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 03:11:33 ; Search time 6386.97 Seconds
(without alignments)
10708.573 Million cell updates/sec

Title: US-10-658-691-4
Perfect score: 1578
Sequence: 1 gtcgacccaatgccatt.....ggcaccggctgtgggtag 1578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_ov.*
22: em_or.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_man.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1578	100.0	6442	1	AY026386	AY026386 Rhodococc
C	228.4	14.5	30600	1	AP005935	AP005935 Bradyrhiz
3	224.8	14.2	1521	6	AR041196	AR041196 Sequence
4	224.8	14.2	1521	6	AR116605	AR116605 Sequence
5	224.8	14.2	1521	6	AR159948	AR159948 Sequence
6	224.8	14.2	1521	1	PSEAMNH	D90216 Pseudomonas
7	218.2	13.8	15355	1	AE005026	AE005026 Halobacte
C	208.4	13.2	308050	1	SC093124	AL939124 Streptomy
9	205.8	13.0	310325	1	AE016864	AE016864 Pseudomon
11	200.6	12.7	1958	1	RERAG	D16207 Rhodococcus
12	187.2	11.9	9330	1	AF315580	AF315580 Agrobacte
13	186.4	11.8	4775	6	E12519	E12519 Nucleotide
C	186.4	11.8	208524	1	AE016925	AE016925 Chromobac
15	182.2	11.5	2822	6	BD061400	BD061400 Rhodococc
16	181.8	11.5	10647	1	AE010321	AE010321 Methanopy
17	180.8	11.5	1879	6	AI9015	AI9015 Brevibacter
18	180.8	11.5	2447	1	BLAMDA	M60264 Brevibacter
19	180.8	11.5	3311	1	RSAMD	X54074 Rhodococcus
20	180.8	11.5	7001	1	AE016078	AE016078 Rhodococc
C	180.6	11.4	180124	1	AE017219	AE017219 Geobacter
22	180.4	11.4	7829	1	RER490527	AJ490527 Rhodococc
23	178.6	11.3	11600	1	AE105912	AE105912 Rhodococc
24	175.4	11.1	1566	6	E12517	E12517 DNA encodin
C	174.2	11.0	300700	1	AP006573	AP006573 Gloeobact
25	172.8	11.0	15939	1	AE007128	AE007128 Mycobacte
C	172.8	11.0	318050	1	EX248344	EX248344 Mycobacte
27	170.4	10.8	1482	6	AX863968	AX863968 Sequence
C	170.4	10.8	348676	15	EX842581	EX842581 Mycobacte
C	170.4	10.8	349726	1	EX640421	EX640421 Bordetell
31	168.4	10.7	346259	1	EX640435	EX640435 Bordetell
32	165.6	10.5	10918	1	AE002025	AE002025 Deinococc
33	158.4	10.0	308650	1	AP005218	AP005218 Corynebac
34	156	9.9	1413	6	BD179870	BD179870 Highly th
35	156	9.9	1416	1	AF202447	AF202447 Thermus t
C	155.2	9.8	110000	2	EX255276_03	Continuation (4 of
C	155.2	9.8	110000	2	EX255276_04	Continuation (5 of
38	154.8	9.8	347365	1	EX569691	EX569691 Synectoco
39	154.2	9.8	346287	1	EX640450	EX640450 Bordetell
C	153	9.7	346897	1	AP002395	AP002395 Mesorhizo
41	152.8	9.7	13698	1	AE005929	AE005929 Caulobact
C	152	9.6	215050	1	AL846057	AL846057 Ralstonia
43	151	9.6	17589	1	AE004862	AE004862 Pseudomon
C	146.4	9.3	12115	1	AE014658	AE014658 Bifidobac
C	146.4	9.3	349980	6	AX492782	AX492782 Sequence

ALIGNMENTS

RESULT 1	AY026386	6442 bp	DNA	linear	BCT 12-JUL-2002
LOCUS	AY026386				
DEFINITION	Rhodococcus erythropolis ISH1 transposase-like protein gene, Partial cds; GntR regulator-like protein, enantioselective amidase (amda), transcription repressor KfrA-like protein, and ISH1 transposase-like protein genes, complete cds; and unknown gene.				
ACCESSION	AY026386				
VERSION	AY026386.1	GI:21734944			
KEYWORDS	Rhodococcus erythropolis				
SOURCE	Rhodococcus erythropolis				
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.				
REFERENCE	1 (bases 1 to 6442)				

AUTHORS	Trott,S., Burger,S., Calaminus,C. and Stolz,A.
TITLE	Cloning and Heterologous Expression of an Enantioselective Amidase from Rhodococcus erythropolis Strain MP50
JOURNAL	Appl. Environ. Microbiol. 68 (7), 3279-3286 (2002)
MEDLINE	22063458
PUBMED	12093004
REFERENCE	2 (bases 1 to 6442)
AUTHORS	Trott,S.
TITLE	Direct Submission
JOURNAL	Submitted (31-JAN-2001) Institut fuer Mikrobiologie, Universitaet Stuttgart, Allmandring 31, Stuttgart 70569, Germany
FEATURES	Location/Qualifiers
source	1..6442
	/organism="Rhodococcus erythropolis"
	/mol_type="genomic DNA"
	/strain="MP50"
	/db_xref="taxon:1833"
	complement (<1..267)
	/codon_start=1
	/transl_table=11
	/product="ISRhl transposase-like protein"
	/protein_id="AAK11723.1"
	/db_xref="GI:21734945"
	/translation="MSTAMPSYKGRHPVEIINHCVLFRPPLSPREVEELMLERGV VSYETIRWCAKFGQAVANQLRRRRPGDRWHLDEVIRINGT"
	complement (1279..2026)
	/note="ORF2"
	/codon_start=1
	/transl_table=11
	/product="Gatr regulator-like protein"
	/protein_id="AAK11723.1"
	/db_xref="GI:21734946"
	/translation="MRPARRAQRVERDLMRSNLLHCAVPDSKLPPEHLMARYDA PRNVEDALALLRHGLDVRVQGYGYALOPPGANDLYDHGVGHPTGTGWSVTRT QVLSHIVTPRAVAGMFGSGERCVCQLDYLAIQGEAVAVATNHFPRCAVVDQP IRTDFELVNRAGIEVGATRFYIGAMDPLHATFQVPENTPVNLSQVIALDGEA IDVAFIWRADRVLLSFVAGPYAGPVGGR"
	2306..3883
gene	/gene="amdA"
	2306..3883
CDS	/gene="amdA"
	/function="enantioselective hydrolysis of ketoprofen amide, naproxen amide, ibuprofen amide, 2-phenylpropionamide, alpha-aminophenylacetamide, alpha-chlorophenylacetamide, alpha-methoxy-phenylacetamide to the corresponding acids"
	/codon_start=1
	/transl_table=11
	/product="enantioselective amidase"
	/protein_id="AAK11724.1"
	/db_xref="GI:21734947"
	/translation="MRPNRFGHVRPTAEQLQXSARHFDLDELAQLVFPVVAEM VTAFDLIDELPQPPTPYTDRIQREPTGDEDPNATFRFCRVEGATEGPLSDITA AIKDCIAAGMPTNGSRMLPTVIATEDAVVVERLLAAGATVGTNLEDMANGIGEG SVYGPALNPNPAHGTGSSGSGAAGVDFALGVDEAGSIRIPRAWCGLVGMKA THGLVPSYGLTYMDHLHDHIGPIITRGVELNARVLEVLADWDQWNLPEPENYG SALGEGSGURFAVSELEPNAGTADVIAANQGLAALASAGATIERVSPVPTAAW PQSGVWAFNARMADESAGVYFHKGRVDVSTAVTTAAQSRTTHKDLAILSLMLVIA EHLDRYGLHYAKAQNRLLELQKQIDVLODRAALLTPTTPTVANELLSGRQDTMSM IPRMTGNAILNTCLDLTGHPLTPTGAGEKGLPVQLQVIGRHFPEESTLYRTGAVIE AAGLWELARPSPAPVLR"
	complement (4171..4740)
	/note="ORF3"
	/codon_start=1
	/transl_table=11
	/product="transcription repressor Kfra-like protein"
	/protein_id="AAK11725.1"
	/db_xref="GI:21734948"
	/translation="MSTSEKIAEAYGVLLARGDKITVRAVOREAGVRIGEVAWMREH AAGAGDVPEAPLSEAVSAMVAVNAWKRAEQADEATAVALDARAGADALVA AEQATBERDGAVTARBAUGEVLRLVEQLRGQDETARQAREARAEESDRARV RAESATDLRLREVLDSLRDNRKPDQPGTS"
	4936..5193
CDS	

repeat_region	repeat_region
CDS	ORIGIN
	Query Match 100.0%; Score 1578; DB 1; Length 6442;
	Best Local Similarity 100.0%; Pred. No. 4.6e-195;
	Matches 1578; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 GTGGACCCCAATGCCCATTCGCCCATGTCCGCCGCCGACAGCCGACGAGCTCCAGGAG 60
DB	2306 GTGGACCCCAATGCCCATTCGCCCATGTCCGCCGCCGACAGCCGACGAGCTCCAGGAG 2365
QY	61 TAGAGCCCGCCACCACTTCGACCTCGACGAGAACTGCCGCCGCCAGCTCGTTCGGGTC 120
DB	2366 TAGAGCCCGCCACCACTTCGACCTCGACGAGAACTGCCGCCGCCAGCTCGTTCGGGTC 2425
QY	121 GTGGCGAGATGGTGACTGCTTCGACCTGATCGAGAACTACCGCAACCGCCGACGCCG 180
DB	2426 GTGGCGAGATGGTGACTGCTTCGACCTGATCGAGAACTACCGCAACCGCCGACGCCG 2485
QY	181 CCGACGCGGTACACGACACCGGACATCGGCCGGGAAACCGACCGCGACGAGACCCGTTT 240
DB	2486 CCGACGCGGTACACGACACCGGACATCGGCCGGGAAACCGACCGCGACGAGACCCGTTT 2545
QY	241 AAGCATTTCATCCGGTTCTCGCGGTGGAGGGGCCACGAGGGGCCACTGAGCGACTG 300
DB	2546 AAGCATTTCATCCGGTTCTCGCGGTGGAGGGGCCACGAGGGGCCACTGAGCGACTG 2605
QY	301 ACCCGCGGATCAAGGACTGATCGCCATCGCCGCTATGCCCATCCACGAAACGGGTCCCG 360
DB	2606 ACCCGCGGATCAAGGACTGATCGCCATCGCCGCTATGCCCATCCACGAAACGGGTCCCG 2665
QY	361 ATGCTCCGACTGTATCGCCACCGAGGATGCCGTGGTGGAGCGGCTGCTCGCGGCA 420
DB	2666 ATGCTCCGACTGTATCGCCACCGAGGATGCCGTGGTGGAGCGGCTGCTCGCGGCA 2725
QY	421 GCGCCACCACTGTCGCAAGACGAACTCGAGGACATGCGGATGGGTATCGGTGAAGGC 480
DB	2726 GCGCCACCACTGTCGCAAGACGAACTCGAGGACATGCGGATGGGTATCGGTGAAGGC 2785
QY	481 AGCGTCTACGTTCTGGCTGAACCGCAACCGCCGCCACGACGGGTGGATCTTCC 540
DB	2786 AGCGTCTACGTTCTGGCTGAACCGCAACCGCCGCCACGACGGGTGGATCTTCC 2845
QY	541 AGCGGCTCCGCGCTGCGCTCGCTGCCGATGTGCTGACCTTCGCCCTGGGCGTCGATGAG 600
DB	2846 AGCGGCTCCGCGCTGCGCTCGCTGCCGATGTGCTGACCTTCGCCCTGGGCGTCGATGAG 2905
QY	601 GCAGGACATCCCGATCCCGCCGATGTGCGGACTGTGCGGATGAAGGCCGCCAC 660
DB	2906 GCAGGACATCCCGATCCCGCCGATGTGCGGACTGTGCGGATGAAGGCCGCCAC 2965

Qy 661 GGCTGTGTCGGCTTACGGCTGACATACATGACACACACCTTGACACATCGGGCC 720
Db 2866 GGCTGTGTCGGCTTACGGCTGACATACATGACACACACCTTGACACATCGGGCC 3025
Qy 721 ATACACAGGGGGTCGAGCTCAAGCCCGGGTCTCTGAGGTGTGGCCGGGGCCGACTGG 780
Db 3026 ATACACAGGGGGTCGAGCTCAAGCCCGGGTCTCTGAGGTGTGGCCGGGGCCGACTGG 3085
Qy 781 CGGACCTCAGTGGGTGCGTAACCTTCGGAGCGGAGAACATACGGCTCCGGCTCGGC 840
Db 3086 CGGACCTCAGTGGGTGCGTAACCTTCGGAGCGGAGAACATACGGCTCCGGCTCGGC 3145
Qy 841 GAGGAGTATCCGGTCTGAGATTCGGGCTCGTGGAGTCACTGGAGCGGAGAACCGTGGC 900
Db 3146 GAGGAGTATCCGGTCTGAGATTCGGGCTCGTGGAGTCACTGGAGCGGAGAACCGTGGC 3205
Qy 901 ACGCCGAGCTGATCGCGCGGTTCACACGAGACTGCGCGCTGAGAGCGCGGTGGC 960
Db 3206 ACGCCGAGCTGATCGCGCGGTTCACACGAGACTGCGCGCTGAGAGCGCGGTGGC 3265
Qy 961 ACCATGAGCGGTCTCGGTGCGTGTGACGCGCGCTGCTATCCAGAGCGCGTG 1020
Db 3266 ACCATGAGCGGTCTCGGTGCGTGTGACGCGCGCTGCTATCCAGAGCGCGTG 3325
Qy 1021 ATGCTTTCAACGCGCGCTATGCGGACTTCCGCGCTTCCGCTTCTCCACAGGGG 1080
Db 3326 ATGCTTTCAACGCGCGCTATGCGGACTTCCGCGCTTCTCCACAGGGG 3385
Qy 1081 CGGTGAGCTGACGCGCGCTCAGACGCGCGCGCGAGTGCACACCGACCGACAGGAT 1140
Db 3386 CGGTGAGCTGACGCGCGCTCAGACGCGCGCGCGAGTGCACACCGACCGACAGGAT 3445
Qy 1141 CTGCGGCTCTGTCGCGGCTGATGCTGCTGATCGCGGACCTGCGGACGATACCTC 1200
Db 3446 CTGCGGCTCTGTCGCGGCTGATGCTGCTGATCGCGGACCTGCGGACGATACCTC 3505
Qy 1201 GGATTCACCTACGCGAGGCGCAGAACCTCGGCTGAGCTCGGCAAGAGATCGAGCC 1260
Db 3506 GGATTCACCTACGCGAGGCGCAGAACCTCGGCTGAGCTCGGCAAGAGATCGAGCC 3565
Qy 1261 GTCTCCAGGACCGGGTGCAGTCTGACCCCGACACCGCTACCGTTCGCAACGAGCTG 1320
Db 3566 GTCTCCAGGACCGGGTGCAGTCTGACCCCGACACCGCTACCGTTCGCAACGAGCTG 3625
Qy 1321 TTGAGCGGTGCGAAGACACATGCTCCATGATCCACGAGTACGGGCAATGCGATCCTC 1380
Db 3626 TTGAGCGGTGCGAAGACACATGCTCCATGATCCACGAGTACGGGCAATGCGATCCTC 3685
Qy 1381 AACACGTGCGCGTGGACCTCACCGGTACCGCGCTGACGCTGCCACGCGTGGCGG 1440
Db 3686 AACACGTGCGCGTGGACCTCACCGGTACCGCGCTGACGCTGCCACGCGTGGCGG 3745
Qy 1441 GAGAGGGCTGCGCGTGGCTCCAGTATAGCGCGCACTTCGAGGAGTTCGAGCTC 1500
Db 3746 GAGAGGGCTGCGCGTGGCTCCAGTATAGCGCGCACTTCGAGGAGTTCGAGCTC 3805
Qy 1501 TACCGACCGCGCGGTGATGAGCGCGCGCTATGGAGCTCGCGCGGAGCGGAGC 1560
Db 3806 TACCGACCGCGCGGTGATGAGCGCGCGCTATGGAGCTCGCGCGGAGCGGAGC 3865
Qy 1561 GCACCGGTGTCGCGTAG 1578
Db 3866 GCACCGGTGTCGCGTAG 3883

RESULT 2
AP005935/c 300600 bp DNA linear BCT 12-SEP-2003
LOCUS Bradyrhizobium japonicum USDA 110 DNA, complete genome, section 1/31.
DEFINITION
ACCESSION AP005935 BA000040
VERSION AP005935.1 GI:27348247

KEYWORDS

SOURCE ORGANISM

REFERENCE AUTHORS

TITLE

JOURNAL MEDLINE PUBMED

REFERENCE AUTHORS

TITLE

JOURNAL MEDLINE PUBMED

REFERENCE

AUTHORS

TITLE JOURNAL

FEATURES

source

gene

CDS

gene

CDS

gene

CDS

```

/codon_start=1
/transl_table=11
/protein_id="BAC45268.1"
/db_xref="GI:27348250"
/translation="MVNAGVVVAPPVGMGRHATGCTGGHIGYRWOIANSVFGLET
QNNADFQGSNANLAFAGVRDSTLDAFLGFTGGVYANNVLLVVRGAAVVRDKYR
VSDFAFLGTDINGSETRMGCTVGAGLEFGFAFNWSIGVEYDHLFMGRDRDVFSPGL
VGVVGTAGTATIGQDVLDIGLVRYNRWGVATY"
4239..4391
/gene="bsr0004"
4391..4391
/gene="bsr0004"
/notes="ORF_ID:bsr0004"
unknown protein
/codon_start=1
/transl_table=11
/protein_id="BAC45269.1"
/db_xref="GI:27348251"
/translation="MGNVDDVMACVIVAADLCARSACRTTPAGADQLGAPRRADAKSR
EVDQR"
complement(4427..4750)
/gene="bll0005"
complement(4427..4750)
/gene="bll0005"
/notes="ORF_ID:bll0005"
unknown protein
/codon_start=1
/transl_table=11
/protein_id="BAC45270.1"
/db_xref="GI:27348252"
/translation="WGLVNHLEVNQRNTATPLSRMSMTATAAPLGLGEGITQTS
EYVKGPNVKGHLHCSPVSGSKPSCSSPFAFDVADPLAVLUGRGAEEVYIEQTL
AWRLG"
complement(4972..6054)
/gene="nrdB"
complement(4972..6054)
/gene="nrdB"
/EC_number="1.17.4.1"
/notes="ORF_ID:bll0006"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase beta
subunit"
/protein_id="BAC45271.1"
/db_xref="GI:27348253"
/translation="MFRMSVIDLSRADPRLIGKRVGLDSTGSDVDRVAMAYEFN
KQQQTHMGEVPLGADNDWASGRVDSREALTIQFRFTQSDIEVGNFYKRYI
PIFQLTQVMNMAFTNMTVHIDAYALLKTLGNPKTEFEAFRGYAGRAKADYMEH
FGVSDVADYARTLAFGATGEMALFASFAMLLNPPHNKMGMOIVSWSVRDESLH
CEGIKLFHEHRETVGATRPVRDDIVDAKTMVKLENFVELAFGLGKIEGMLPEHI
HAYRVADWRLAQLRIAPVFGFEAKESGGFTQLKAFLPLWLVILNGVHEANFEQR
ATEYSKASRGSWDGEDGVWATFGRI"
complement(6036..8207)
/gene="bll0007"
complement(6036..8207)
/gene="bll0007"
/notes="ORF_ID:bll0007"
probable ribonucleoside-diphosphate reductase (EC
1.17.4.1)
/codon_start=1
/transl_table=11
/protein_id="BAC45272.1"
/db_xref="GI:27348254"
/translation="MLGNPLCRQLFRGISATALGRPTAAATREPGLNPRNPCPRA
GCPVGRLOFPARPKRSARSOPNITGLPMSASYDFDOYGHLELPSVROPSSAQA
KPTTELICIVOPPALPRLPPAPBSPTPSDLSDRSDDLTSFGRLTLTDVILLAG
EGLQMFARVSCAFADNVAHQRLYDMSRLWFNPATPVLNNGTTLGLPLSCFNSV
SSLSGIVGTNENWLASNGGIGTYGNVRSIGKARGVGTGIIPIPTHVDTGLTL
ALSQSLRGSNTNEVRIEHLHFEIEFLIRKASGDFNRKSLNHHGINVTFDEFNCAVG
ACTALRSPKTEVRIEDARLQWRIENRLOTGEBPYLLFIDTVNRLPKHQRELG
LEVSTNSLCEITLSTGDHGRGERTAVCCLSLNLETDOWNEEPGEFIEDVMEFLDN
VLTHITVAPDGMKARVAALRSRVGLGVMGHFSFLOAKGIPMEGLAKSFNLMFR
KLRRADAASVALARGPCPDALERGVNARFESHKIAIAPTASISICGGTSACTEPI
```

```

PANIYTHKLSGALSVRNPHLAKVLAAGKADTSATWESTIAHDSGVAHLDVLSHEHKA
IFRTAFEDQRIWVLEADRAALICQSINLYLPADVDKWDLHMLHWTAMKRGVKSJ
YYCRSISRAAPAGKIAGDATPTQPARQRTDYERCLAQ"
complement(8380..8688)
/gene="bll0008"
complement(8380..8688)
/gene="bll0008"
/notes="ORF_ID:bll0008"
unknown protein
/codon_start=1
/transl_table=11
/protein_id="BAC45273.1"
/db_xref="GI:27348255"
/translation="MARRYWLKPKDPAQLNSRTANRSMPTFRMSRVAVLLENLRHFP
ORFHLADRRRVVVERTFANLGRCHLCKDFEGSAATLAWPLVAHLRLTLRLARP"
complement(8962..9714)
/gene="bll0009"
complement(8962..9714)
/gene="bll0009"
/notes="ORF_ID:bll0009"
putative acetyltransferase
/codon_start=1
/transl_table=11
/protein_id="BAC45274.1"
/db_xref="GI:27348256"
/translation="MHPASPRRCIHSARQMCRCCHAPPKANAKTEPRNDFMTBESDI
RSANANTRASTVTVIVLGRADPLNWSWPDPROVLSRMPQFIDACGAFAPAGTAYV
AGIRGALWLPVNVQDDEALDAINALSIRFEITDMAHRRMAHEHPHEPHWLP
LIAVDNWKVGKGLTLMNYALUCRDEGDGIVAYLESSNPENPFYQRHGKFKVVGELQH
GSPPFTMLRTASFSPSKAAPDRLVTQQA"
complement(10345..10743)
/gene="bll0010"
complement(10345..10743)
/gene="bll0010"
/notes="ORF_ID:bll0010"
unknown protein
/codon_start=1
/transl_table=11
/protein_id="BAC45275.1"
/db_xref="GI:27348257"
/translation="MLPRLOFSKTERNNTHVDLPGVVASIPAPRNLNRSDOGV
WNSYRAINACGOKPLRLRLQLRQPAVFPPTMARIAATTLGLGVYDEGTTATFSLD
RANAVRADTDPDQSMHESRAIPSCVVQ"
complement(10872..11492)
/gene="bll0011"
complement(10872..11492)
/gene="bll0011"
/notes="ORF_ID:bll0011"

Query Match 14.5%; Score 228.4; DB 1; Length 300600;
Best Local Similarity 50.9%; Pred. NO. 4.1e-21;
Matches 677; Conservative 0; Mismatches 636; Indels 17; Gaps 5;

Qy 217 CCGACGGCGGACGAAGACCCGTTCAACGGCATTCATCCGGTTCTTCCGGGTGAGGCGGCC 276
Db 194385 CCCTCGGCGAAGAGAACCCGCGCATGCTGTGATCGCAAGCGCATGTGAAAGCGCC 194386
Qy 277 ACGAGAGGGCCATGAGCGACCTGACCGCGCGGATCAAGACTGATCGCCATCGCCGGT 336
Db 194325 GCCGCGGCAAGCTCAAGGGCAAAACCGTTGCGCTGAAAACAACTCATGTGCGCCGC 194266
Qy 337 ATGCCACACCAAGACCGGTCGCCGATGCTCCCGACTGTGATCGCCACCGAGGATGCCGTG 396
Db 194265 GTGCCCATGACCAACCGGCTGCTGCGAGGCTATGTGCCCATTCGACGCCACC 194206
Qy 397 GTGGTGAGCGGTGCTCGCGGAGCGGCCACCATCGTCGGCAAGACGAACCTCGAGGAC 456
Db 194205 ATCGTCACGGCATGCTGGATGCGGTGCGGAGATCAAGGGCAAGGTGATCGCAGCAT 194146
Qy 457 ATGCGCATG---GGTATCGGTGAGGACGCTACGGTCTCGCGCTGACCGGACCAAC 513
Db 194145 TTCTGCTGTCCGCGGCGCACACCGGCTCTTTTCGGGCCCGTGCAATTCGCCAAG 194086
```


QY	514	CCCGCCCAACGGCA	CGGGTGGATCTT	CAGGGGTCTCGGGCTCCGTCCTCGCTCCGCGCATG	573
Db	194085	ATGGGTACT	TCGGCCGGTGGCT	CTGCTGTCAAGGCTCGGGGCTGCTGGTCTCGCTCCGGCGAA	194026
QY	574	GTGCACTTCGGCCCT	GGGCGTTCGAT	TAGGCGAGGAGCATCCGGATCCCGCCGCACTGGTGC	633
Db	194025	GTGCACATGGCGAT	TCGGCGCGCAT	CAGGGCGGTTCGATCCGATCGATCGCTCCGTCCTCGTTCGTC	193966
QY	634	GGACTGGTCGCATGA	AGGCGACCA	CGGCGCTGGTGCCTTACCGGCTGACATACATG	693
Db	193965	GGCAATTAGGCA	TGAAGCCACT	TGGGCGCTCTGTCCTTACACCGGATCATGCGGATC	193906
QY	694	GACCACACTTTGG	ACCAATCCGGGCCCAT	CACAGGGGGGTGAGTCAACGCCCGGCT	753
Db	193905	GAGATCTATATCG	ACCATACCCGGTCCGAT	CACCGCACCTCGCTGCATCAACCGCTGCTG	193846
QY	754	CTCGAGGTGTTGG	CGCGGGCGCACT	TGGCGGACCTTCAGTGGGTGCGTAACCTTCGGAG	813
Db	193845	CTTGAAGTGTCT	CCCGCGCATG	ACGGCTACGATC-----GGCATCAAGGCTCGAAG	193792
QY	814	CCGGAGAACTAC	GGCTCCGCGCT	CGGGAGGAGTATCCGGTCTGAGATTCGGGTGCTGTC	873
Db	193791	GTCCAGGAATA	CACCAAGCGCT	CGGCAGGGCTCAAGGGCATGAAGATCGCATCTCTC	193732
QY	874	GAGGAGTCACTG	SAGSCGAA	CGGTGCGACGCGGACGTGATCCCGGCTTCAACACAGGA	933
Db	193731	AAAGAAAGCTTT	TGACAGCGCGGT	TCGAGAAGCGCGGTGATGAAGCGTCCGAGGGC	193672
QY	934	CTGGCGGCGCT	TCGAGAGCGCGGT	TCGAGACATCGAGCGGTCTCGGTGCGTGTGGAGCG	993
Db	193671	GCAAGCGCTTCA	AGGATCTTTGGCGCC	ACCGTAGAGCGGTCTCAATCCCGATCGATCTC	193612
QY	994	GGGCGCTGCGCT	TATCCAGAGCGCGT	GATGCTTCAACGCGGCGCTATGGCGGACTCC	1053
Db	193611	CTGGGCGGCGAT	CTGGAACCCGAT	CGGCAACGAGCCCTACCCAGACCATGATGTC	193552
QY	1054	GCGGTGTGGGTA	CTTCCAAGGGGGCG	GTGAGCTGACACCGCGCTCACGAGCGG	1113
Db	193551	GCGCAGCGGTAC	GGGCTCAGCGCGCGAT	CTCTATTTCGACCTCGGTGATGGATTTTCAT	193492
QY	1114	GCCAGAGTCGCA	CCACCAAGGAT	CTGGCGATCTGTCCGGCTGATGCTGGTGATC	1173
Db	193491	CGCGGTGCGCG	CGGACGGCACTCG	GTGTCCGAGCCACGAGCTGTTCATGATGCTC	193432
QY	1174	GCGGAGCACTCG	GGCGA	CGAATACCTCGGATCCACTACGGGAAGGCGCAGACCTGCGG	1233
Db	193431	GGCACCTACAT	CAACAACAAATTC	CGGTCCACGCTTCTACGCGAGGCGCTCAATATCTCC	193372
QY	1234	CTGGAGCTCGGA	AGAGATCGACCCGCT	CTCCAGGACCGGGTGCACCTGCTGACCCG	1293
Db	193371	CGCAGCTAC	CGCGCTACGAC	AAAGCGTTTGGGATTAAGATCTGCTTTTGTGCGCG	193312
QY	1294	ACCACGCTACCT	GTGCCAA	CGAGTGTGAGCGGTTCGGCAGACACCATGTCCATGAT	1352
Db	193311	ACGACCCGATGA	AGCAACCAAGCT	TCGGAGCCGACCGCAGGAGTACGTC	193252
QY	1353	CCACCGATGAC	GGGCAATGCGAT	CCTCAACACGTGCCGCTGAGACCTCACCGGTCAACC	1412
Db	193251	GCCCGTG----	CGCTGGAGAT	CGGCAACACCGCGCGTTTCGACATCACCATCACCC	193196
QY	1413	GGCGCTGAC	GTGTGCCCA	CGGTTGCGGGCGAGAGGGGCTGCGCTTGGCTTCCAGTGAT	1472
Db	193195	CGCGATGT	CGCTGCCCT	CTGGGATGTCG-----ACGGCTGCGCTCGGCTGATGTGGT	193139
QY	1473	AGGCGCGCACT	TCGAGGAGT	CGACGCTTACCGCACCGGCGCGTGTACGAGCGCGCG	1532
Db	193138	CGGCGCGAT	TTTCGAGGAT	CGACCATCTATCGCGCGCGCACCGCTTCGAGCAGATCGG	193079
QY	1533	CTTATGGGAG	1542		
Db	193078	CGACTGGA	193069		

RESULT 3	AR041196	1521 bp	DNA	linear	PAT 29-SEP-1999
LOCUS	Sequence 20 from patent US 5811286.				
DEFINITION	AR041196				
ACCESSION	AR041196				
VERSION	AR041196.1	GI:5961692			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
1 (bases 1 to 1521)					
REFERENCE	Fallon,R.Donald., Nelson,M.James. and Payne,M.Scott.				
AUTHORS	Nucleic acid fragments encoding stereospecific nitrile hydratase				
TITLE	and amidase enzymes and recombinant organisms expressing those				
	enzymes useful for the production of chiral amides and acids				
JOURNAL	Patent: US 5811286-A 20 22-SEP-1998;				
FEATURES	Location/Qualifiers				
source	1..1521				
	/organism="unknown"				
	/mol_type="unassigned DNA"				
ORIGIN					
	Query Match 14.2%; Score 224.8; DB 6; Length 1521;				
	Best Local Similarity 48.8%; Pred. No. 5.5e-20;				
	Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;				
Qy	37	CCGACACCCGAA	CAGCTCCAGGAGTACAGCGCCGCCACCACTTCGACCTCGACGAGAA	96	
Db	16	CCTACCTTCGAC	CAAGGTTTGTAGCATCCGAACCCAGTTGCACATGCAACTGACGACGAA	75	
Qy	97	CTGCGCCGCCAG	CTCGTTCGGTCTGTCGGGAGATGTGTGACTGCTTTCGACCTGATCGAC	156	
Db	76	CAGGACGCGTCT	ACCTTGGAACTGATGCAACCCAGAGTTTCGACGCTACGACCTGGTCGAC	135	
Qy	157	GAACTACGCA	ACCGCGACGCCCGCGGTACACGGACCGACATCGGCCGCGAA	216	
Db	136	GAACTGGCTGAT	TTTCGTCGCCAATACGCTACGACCGCATTCAGGCTATCGCCATCGG	195	
Qy	217	CCGACCCGGCAG	GAAGACCCGTTCAACGCAATTCATCCGGTTCGCCGGGTGGAGGGCGCC	276	
Db	196	CCATCGCGCAAG	GAAGAAACCTCTGAAAGCCTGTGTACTACCGAACAGAAAGTGAATGTCGC	255	
Qy	277	ACGAGAGGGC	CACTAGCGACCTGACCGCGCGATCAAGGACTGCATCGCATCGCCGGT	336	
Db	256	CGGAAAGCCTTG	CTGGCGGGCAAAACGTCGCGCTCAAGATAATATCTCCCTGCGAGGC	315	
Qy	337	ATGCCCAACCA	CGAACCGGTCCGGATGCTCCGCACTGTGATCGCCACCGAGGATGCCGTG	396	
Db	316	GTCCCCATGAT	GAACCGCGCAGCGCCGTTGGAAAGGCTTCGTCCCGGGGTTCGATGCCACG	375	
Qy	397	GTGTGGAGCG	GCTCTCGCGCAGGCGCCACCACTCGTCGGCAAGACGAACCTCGAGGAC	456	
Db	376	GTGTTCACCC	CTTGCTCGATCGCGGGCGACCAATTCGCGCAAGCACTCGCGAGCAC	435	
Qy	457	ATGGCGAT- -	-GGTATCGGTGAAGGAGCGGTCTACGGTCTGTCGCTGAAACCCGAAACAC	513	
Db	436	TACTGCTTTT	CAGGAGGCAGCACACCTCCGATCCAGCCCGCGTGCACAAACCCACATCGC	495	
Qy	514	CCGCCACACG	CACGGGTGGATCTTCAGCGGCTCCCGCGCTGCGCTCGCTCCCGCGCATG	573	
Db	496	CACGGTTATG	CTTGCGGTTCTTCATCAGGACGCGGGCATTTGTTGCTCCGGTGAG	555	
Qy	574	GTCGACTTCG	CCCTGGGGCTTCGATGAGGCAAGGACGATCCGGATCCCGGCCGCAATGGTGC	633	
Db	556	GTGGACATCG	CGTGGCGGCGATCAAGGCGGCTCCATTCGGATCCCGCTCGGCCCTTCGC	615	
Qy	634	GGACTGGTGC	GATGAAGGCGACCCACCGGCTGGTGGCGCTTACGGGCTCAGACATACATG	693	
Db	616	GGTACCTTAG	CGATGAAGCCACCAACCGGCTTGTGCTCCTACACGCGGCTCATGGGATT	675	
Qy	694	GACCAACCTT	TGGACCATTCGGGCCCATCAACGAGGGGGTTCGAGCTCAACGCCCGGGTGC	753	

Db 676 GAAGCCACGATCGATCATGTGCGGCCCATACCGGTAAAGTGGCGCAACACGGCGTGATG 735
Qy 754 CTCGAGGTGTTGGCCGGGCGGAGTGGCGGACCCCTCAAGTGGGTGCGTAACCTTCCGGAG 813
Db 736 CTCGAGGCAATGGCGGTGCGAGAGGACTCGACCCGCGCCAGCGCGCG-CTTCAG 789
Qy 814 CCGGAGAACTACGGCTCCCGCTGCGCGAGGAGTATCGGTCTGAGATTGCGGTGCTC 873
Db 790 GTCGATGACTATTGAGTTACTCTGAAAGAGCGGTGAGCGGACTCAGAACTCGGGTGTG 849
Qy 874 GAGGAGTCACTGAGCGGAAACGGTTCGACGCGGACGTGATCGCGCGCTTCAACACAGGGA 933
Db 850 CAAGAGGGATTGCGCTTCTTAACAGGACCCCTCGCGTGGGGAACAAAGTGGCGACGCC 909
Qy 934 CTGGCGGCGCTGAGAGCGCGGTGCGACCATCGAGCGGTCTCGGTGCGGTGTTGGAGC 993
Db 910 ATCGCCCGACTCGAGCGGTGGCGCTCATGTGAGCGCGTCTCATTCGCGAGCAAC 969
Qy 994 GCGGCTGCGCTATCCAGAGCGCGCTGATGGCTTTCAACGCGCGCGCTATGCGCGACTCC 1053
Db 970 CTGGCAGGGTGTGTTGGCACCCCATCGGTTCGGAAGGCTTGACCATGCAGATGATCAT 1029
Qy 1054 GCGGCTGCGGTACTTCCAAAGGGCGCGTGGAGCTGAGCAGCGCGTTCAGACGGGG 1113
Db 1030 GGAACGCGCGAGGCTTTAACTGGAAAGGACTTTACGATGTGCGGCTGCTGACAAACAA 1089
Qy 1114 GCGGAGTCCGACACCAACGATCTGGGATCTGGCGATCTGTCGCGCTGATGCTGTGATC 1173
Db 1090 GCGAGTGGCGGACGACGACGACCAATTTATCGCGTGCCTCAAGCTCTGATGTTGCTC 1149
Qy 1174 GCGGAGACTCTGCGGACGAAATACCTCGGATCTCACTACGGAAGCGCGAGAACTCGGG 1233
Db 1150 GGCACATAGCGCTGTGCGGTACAAACGAGCGCTACTACGCAAGCGCGCGAGAACTTGA 1209
Qy 1234 CTGGAGTCCGCAAGCAGATCGACGCGCTCTCCAGGACCGGCTGCACTGTGACCGCG 1293
Db 1210 CGTTTTCGCGGACGAGTACGACAAAGCGCTGCAACCTATGACTGTGATGTCGCG 1269
Qy 1294 ACCAGCTACCTGTCGCAAGAGCTGTTGAGCGGTGCGGAGACACCAATGTCATGATC 1353
Db 1270 ACCAGCGCATACGCGCCCAACCCACCGCGCAG---CGAAGTGTGCTGATCGAGTAC 1326
Qy 1354 CCAGGATGACGGGCAATGATCTCAACAGTGGCGCTGCACTCAACCGGTCAACCG 1413
Db 1327 GTGGCTCGCGGTGGAAATGATCGGCAATACCGCGCCACAGACATCACCGGATCGG 1386
Qy 1414 GCGCTGAGCGTCCGCGGAGGAGGCGCTGCGGTGGCGGAGGCGCTTCCGCTTCCAAAGTGA 1473
Db 1387 GCCATGTGATTC---CGTGGCTGCTGGAGCGGCTGCGCGTGGGTGATGCTGCTGTC 1443
Qy 1474 GCGCGCACTTCGAGGAGTGCAGCTTACCGCACCGGCGCGCTGATCGAGGCGCGC 1531
Db 1444 GCAAAACACTACGCGGAGGCGATTTACCAAGCGCGCGCGGCTTGAAGCGCTCGG 1501

RESULT 4
ARL16605
LOCUS ARL16605 1521 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 20 from patent US 6133421.
ACCESSION ARL16605
VERSION ARL16605.1 GI:14096927
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1521)
AUTHORS Fallon,R.Donald., Nelson,M.James. and Payne,M.Scott.
TITLE Polypeptides and polypeptide subunits of a stereospecific nitrile
hydratase enzyme
JOURNAL Patent: US 6133421-A 20 17-OCT-2000;
FEATURES Location/Qualifiers
source 1..1521
/organism="unknown."

ORIGIN
Query Match 14.2%; Score 224.8; DB 6; Length 1521;
Best Local Similarity 48.8%; Pred. No. 5.5e-20;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;
/mol_type="unassigned DNA"
Qy 37 CCGACAGCGAAGCAGCTCCAGGAGTACAGCGCCCGCCACACTTCGACCTCGACGAGGAA 96
Db 16 CTTACCTCTACCAAGTCTTTAGACATCCGAACCCAGTTGCAATGCAATGACGACGAA 75
Qy 97 CTGGCGGCGGAGCTCGTTCCGCTGCTGGCGGAGATGCTGCTTCGACCTTCGACCTGATCGAC 156
Db 76 CAGGAGCGCTCTACCTGGAACTGATGCAACCGAGTTTCGACGCTTACGACCTGCTGAC 135
Qy 157 GAATACCGCAACCGCGGACCGCGCGCTACACGCGCGGACATCGCGCGGAA 216
Db 136 GAATGCTGCTATTTGCTTCGCGCAATACGCTAGCAGTTCAGGCTATCGCCATCGG 195
Qy 217 CCGACCGCGGACCAAGAACCGGTTCAACGCAATTCATCCGTTTCGCGGGTGGAGGCGCC 276
Db 196 CCATCGGCAAGGAAACCTCTGAAAGCTGTGTTACTACCGAACAGAGTGAATGCTGCC 255
Qy 277 ACGAGGGGCGACTGAGCGACCTGACGCGCGGATCAAGGACTGCTTCGCCATCGCGGT 336
Db 256 CGGAAAGGCTGTGCGGGGCAAAACGTCGCGCTCAAGATAATATCTCCCTGGGAGGC 315
Qy 337 ATGCCACCAACGAGTCCCGATGCTCCGACTGTGATCGCCACCGAGGATGCGGTG 396
Db 316 GTCCCATGATGAACGGCGCAGCGCTTGGAGGCTTCGTCCCGGGTTTCGATGCCACG 375
Qy 397 GTGTGAGCGGTGCTGCGGCGAGGCGCACCATGTCGCGGAAGAGAACTTCGAGGAC 456
Db 376 GTGTTCACCGGCTGTGCTGATGCGGGGCGACCATTTCTCGGCAAGGCGACCTCGGAGC 435
Qy 457 ATGGCGAT---GGGTATCGGTGAAGGCGAGCTGTACGCTCTCGCTGAAACCCGAAAC 513
Db 436 TACTGCTTTGAGAGGAGCGCACCATCTCCGATCCAGCCCGGTGCAACACCATCGC 495
Qy 514 CCGCGCCACGCGACGCGGTGATTTCCAGCGGTTCGCGCGCTGCGCTGCTGCGCGCATG 573
Db 496 CACGTTATGCTCTGCGGTTCTCATACGAGCGCGCGCATTTGTTGCTCGGTGAG 555
Qy 574 GTCCACTTCCGCTGCGGTGCTGATGAGGAGGAGCATCCGATCCCGCGCGCATGCTGC 633
Db 556 GTGACATCGCGCTGCGGCGGATCAAGGCGCTCATTCGATCCGCTCGGCTTCTGC 615
Qy 634 GACTGCTGCGCATGAAGGCGACCCACGCGCTGCTGCTGCTTTACGCGCTGACATCATG 693
Db 616 GGTACTAGCGCATGAAGCCACCCACGCGCTGCTGCTGCTTACACCGCGCTCATGGCGATT 675
Qy 694 GACCACTTGGACCATCGGCGCCATCACAGGCGGCTGCGAGCTCAACGCGCGGCTC 753
Db 676 GAAGCAGATCGATCATGTGCGGCCCATCACCGTTAACGTTGCGCGCAACGCGCTGATG 735
Qy 754 CTCGAGTGTGTCGCGGCGGCGACTGCGCGGACCCCTCAGTGGGTGCGTAACCTTCCGGAG 813
Db 736 CTGCGGCAATGCGCGGTGCGAGCGGACTCGACCCGCGCGAGCGCGG-----CCTCAG 789
Qy 814 CCGGAGAACTACGCTCGCGCTCGGCGAGGAGTATCCGCTCTGAGATTGCGGTGCTC 873
Db 790 GTCGATGACTATTGAGTTACTCTGAAAGAGCGGTGAGCGGACTCAGAACTCGGGTGTG 849
Qy 874 GAGGAGTCACTGAGCGGAAACGGTTCGACGCGGACGTGATCGCGCGGTTCAACACAGGGA 933
Db 850 CAAGAGGATTCGCGCTTGTCTAACAGGACCCCTCGCGTGGCGGACAAAGTGGCGACGCC 909
Qy 934 CTGGCGGCGCTCGAGAGCGCGGTGCGACCATTCAGCGGGTCTCGGTGCGGTGTTGAGC 993
Db 910 ATCGCCCGACTCGAGGCGTGGCGGCTCATGTGAGCGCGTCTCATTCGCGAGCAAC 969
Qy 994 GCGGCTGCGCTATCCAGAGCGGCTGATGCTTTCAACGCGCGCGCTATGCGCGACTCC 1053

Db 970 CTGGCAGGTTGTGTGGCACCCCATCGGTTGGAAGGCTTGACCATGCGAGTATGTCAT 1029
Qy 1054 GCCGGTGTGGGCTACTTCCACAAGGCGCGCTGGAGCTGAGCAGCAGCCGCTACGAGCGG 1113
Db 1030 GGCAACGGCGCAGGCTTTACTGGAAGGACTTTACGATGTGGCTGTGACAAACA 1089
Qy 1114 GCCAGAGTCGACACCCACAAGGATCTGGCGATCTCTGCGGCTGATGTGTGATC 1173
Db 1090 GCCAGCTGGCGCAGCAGCAGCAACCAATATCGCGCTGCTCAAGCTCTGCTATGTCGTC 1149
Qy 1174 GCGGAGCAGCTGCGCAGCAATACCTCGGATCCACTACGGAAGCGCAGAACCTGCGG 1233
Db 1150 GGCAATACGGCTGTGCGCTACAGAGCGTACTACGCCAAGGCCAGAACCTTGCA 1209
Qy 1234 CTGGAGCTGGCAAGCAGATCGAGCGCTCTCCAGGACCGGCTGCACTGTGACCCCG 1293
Db 1210 CGCTTTTCCCGCAGGAGTACGACAAGCGCTGCAAACTATGACCTGTGCTGTGTCGCG 1269
Qy 1294 ACCACGCTACCGTTGCCAAGAGCTGTGAGGCTGGCGGTCGCAAGACACCATGTCATGTC 1353
Db 1270 ACCACGCCCATACCGGCCCAACCCCAACCGCCAG---CGAATGCTGCTATCGAGTAC 1326
Qy 1354 CCACGAGTACGCGCAATCGATCTCAACACGCTGCGCGCTGACCTACCGCTCACCGG 1413
Db 1327 GTGGCTCGCGCTGTGAATATGTCGCAATACCGGCCACAGACATCACCGGCTACG 1386
Qy 1414 GCGCTGACGCTGCGCAGGCTGGCGGCGAGAGGCTGCGCTGCGCTCCAAAGTATA 1473
Db 1387 GCCATGTGCAATC---CGTGTGCTGCTGAGCGGCTGCGCTGCGCTGCTGCTGTC 1443
Qy 1474 GCGCGGCACTTCGAGGAGTCAAGCTCTACCGCAGCGCGGCTGATGAGCGCGCGG 1531
Db 1444 GCAAAACACTACCGCAGGCGCAGATTACCAAGCGCGCGGCTGTTGAGCCTCGG 1501

RESULT 5
AR159948
LOCUS AR159948 1521 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 20 from patent US 6251650.
ACCESSION AR159948
VERSION AR159948.1 GI:16222813
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1521)
AUTHORS Fallon,R.Donald., Nelson,M.James, and Payne,M.Scott.
TITLE Pseudomonas putida amidease polypeptide useful for the production of
chiral amides and acids
JOURNAL Patent: US 6251650-A 20 26-JUN-2001;
FEATURES
source
1. .1521
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match 14.2%; Score 224.8; DB 6; Length 1521;
Best Local Similarity 48.8%; Pred.No.5.5e-20;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;
Qy 37 CCGACAGCGCAAGCTCCAGGAGTACAGCGCGCCGACCTTCACCTCCGACGAGAA 96
Db 16 CCTACCTCGACAGGTTTAGACATCCGAACCACTTGCAATGCACTGACGACGAA 75
Qy 97 CTGGCGCCAGCTCGTTCCGTCGTCGGGAGATGCTGCTGCTGACCTGATCGAC 156
Db 76 CAGGAGCGTCTTACCTGGAATGATGCAACCGAGTTTCAAGCGCTTACGACCTGCTGAC 135
Qy 157 GAACTACCGCAACCGCGGAGCGCGCAGCGCTTACAGGACCGGACATCGCGCGAA 216
Db 136 GAACTGCTGATTGTTGTTGCGCCAAATACGCTAGCAGCGGAGTTTCAAGCTATCGCCATCGG 195
Qy 217 CCGACCGCGCAGGAACCGGTTCAAGCATTCATCCGCTTCTCGCGGTTGAGGCGCGC 276

196 CCATCGGCAAGGAAACCCCTCTGAAACGCTGGTACTACCGAACAGAAAGTGAATGGTGCC 255
277 ACAGAGGCGCAGTGGAGGACTGACCGCGGATGAGGACTGCAATGCGCATCGCCGCT 336
256 CGCAAGGCGCTGCTGGCGGCAAAACCGTGGCGCTCAAGATATATATCTCCCTCGCAGGC 315
337 ATGCCCAACCAAGGAGTCCCGGATGCTCCGAGTGTGATCGCAGCAGGAGTCCGCTG 396
316 GTCCCATGATGAACGCGGAGCGCGCTTGAAGGCTTCGTCGCGGTTGATGCCACG 375
397 GTGCTGAGCGGCTGCTGCGGAGGCGGACACCATCTGTCGCAAGAGCAAGCTCGAGAC 456
376 GTGCTACCGGCTGCTGCTGATGCGGCGGCGACCAATCTCGGCAAGCACTCGGAGCAC 435
457 ATGGCGAT---GGGTATCGGTGAAGGAGGAGGCTTACCGTCTGCGCTGAACCCGAAAC 513
436 TACTGCGCTTTCAGGAGGAGCACACCTCCGATCCAGCCCGGCTGACACCCACATCGC 495
514 CCGGCCCAAGGCAAGGCTGATCTTCAGCGGCTTCGCGGCTGCGCTGCTGCGGCGATG 573
496 CACGTTATGCTCTGCGGCTTCCTCATAGGAGGCGCGCATTTGGTTGCTGCGCTGAG 555
574 GTGACTTTCGCTGCGGCTGATGAGGAGGAGCATCGGATCCCGGCGCATGCTGTC 633
556 GTGACATCGCTGCGGCGGATCAAGGCGCTCAATTCGATCCGCTGCGCTTCTGC 615
634 GGAATGTGCGCATGAAGGAGCAGCGGCTGCTGCGCTTACGCGCTGACATACATG 693
616 GGTACTACGCGCATGAAGGCGCAGCGGCTGCTGCGCTTACACCGGCTGATGCGCAT 675
694 GACCACACTTGAACCATCGGCGGCTTACAGGAGGAGGCTGAGCTCAAGCGCGGCTG 753
676 GAGCCAGATCATGTCGCGGCTTACCGGCTTACCGGTAACGTCGCGGACACCGCTGATG 735
754 CTCGAGGTTTGGCGGCGGCTGCGGAGCCTCACTGAGTGGTGGTGTACCTTCGCGAG 813
736 CTGAGGCAATGCGGCTGCGGAGGAGTATCGGCTGCTGAGTTCGCGCTGCTG 789
814 CCGGAGGACTACGCTCGGCTGCGGAGGAGTATCGGCTGCTGAGTTCGCGCTGCTG 873
790 GTCGATGACTATTGCACTTCTGGAAGGCGTGAGCGGCTCAGATTCGCGGCTGTTG 849
874 GAGGAGTCACTGGAGCGGAGCGGTGCGAGCGCGGAGCTGATCGCGGCTTCAACAGGGA 933
850 CAAGAGGATTCGCGCTTGTCTAACAGGAGCTTCGCGTGGCGGAGAAAGTGGCGAGCC 909
934 CTGGCGGCTGCGAGGCGGCTGCGAGCCTACGAGGCGGCTGCTGCGCTGCTGAGACG 993
910 ATCGCCGAGCTGAGGCGGCTTGGCGCTCATGTGAGCGCGCTCTCATTCGCGAGCAAC 969
994 GCGGCTGCGCTATCCAGAGCGGCTGATGGCTTCAACGCGCGCGCTATGGCGGAGTCC 1053
970 CTGGCAGGTTGTTGTGCAACCCCATCGTTGCGAAGGCTTGACCATGCGAGTATGAT 1029
1054 GCGGCTGCGGCTATCTTCCACAAGGCGGCTGAGAGCTGAGCAGCGCGCTCAGAGCGG 1113
1030 GCGAACGCGCGAGGCTTTAACTGGAAGGAGCTTTACGATGTCGCGCTGCTGGAACAA 1089
1114 GCCCAGAGTGCACACCAAGGATCTGCGATCTGCTCCGCGCTGATGCTGCTGATC 1173
1090 GCGAGCTGCGCGAGCGAGCAATTTATCCGCTGCTGCAAGCTTCTGATGTTGCTG 1149
1174 GCGGAGCACTGCGGAGCAATACCTCGGATCCACTACGGAAGGCGGAGAACTCGGG 1233
1150 GCGCAATACGCGCTGTGCGCTACAGGAGCGCTTACTACGCAAGGCGGCGGAGAACTTGA 1209
1234 CTGGAGCTGCGCAAGAGATGAGCGGCTCTCCAGGAGCGGCTGCTGCTGCTGCTGCTG 1293
1210 CGCTTTCGCGGAGGAGTACGAAAGCGCTGCAAACTATGATGCTGCTGCTGCTGCTG 1269
1294 ACCACGCTACCGTTGCGCAACGAGCTGTGAGCGGCTGCGGAGCAACCATGTCATGATC 1353

Db 1270 ACCACGCCATCAGCGCCCAACCCACCCGCCAG---CGAACTGCTCGATCAGCGAGTAC 1326
Qy 1354 CCACGATGACGGCAATCGCATCTCAACAGCTGCCCGCTGGAGCTCAGCGGTCAACCG 1413
Db 1327 GTGGCTCGCGCTTGGAAATGATCGGCAATACCGGCCACAGGACATCACCAGGATCCG 1386
Qy 1414 GGGCTGACGTCGCCACCGGCTCGCGGCGAGAGGCGCTGCCGCTTGGCTCCAGTGAATA 1473
Db 1387 GCGATGTCGATTC---CGTGTGGCTGCTGACGCGCTGCCGCTCGGCTGATGCTGTC 1443
Qy 1474 GCGCGCCACTTCGAGGAGTCGACGCTTACGCGACCGCGCGGCTGATCGAGCGCGCG 1531
Db 1444 GCARAACTAGCCGCGAGGACGATTTACAGCGCGCGCGGCTTTGAAGCTCGG 1501

RESULT 6
LOCUS PSEAMNH 5215 bp DNA linear BCT 21-DEC-2002
DEFINITION Pseudomonas chlororaphis genes for amidase and for nitrile
ACCESSION D90216 M65204
VERSION 1
KEYWORDS D90216.1 GI:216850
Qy 1474 GCGCGCCACTTCGAGGAGTCGACGCTTACGCGACCGCGCGGCTGATCGAGCGCGCG 1531
Db 1444 GCARAACTAGCCGCGAGGACGATTTACAGCGCGCGCGGCTTTGAAGCTCGG 1501

ORGANISM Pseudomonas chlororaphis
SOURCE Pseudomonas chlororaphis
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 5215)
AUTHORS Nishiyama,M., Horinouchi,S., Kobayashi,M., Nagasawa,T., Yamada,H.
and Beppu,T.
TITLE Cloning and characterization of genes responsible for metabolism of
nitrile compounds from Pseudomonas chlororaphis B23
JOURNAL J. Bacteriol. 173 (8), 2465-2472 (1991)
MEDLINE 91193202
PubMed 2013568

COMMENT These data kindly submitted in computer readable form by: Makoto
Nishiyama
Department of Agricultural Chemistry
The University of Tokyo
1-1-1 Yayoi, Bunkyo-ku
Tokyo 113
Japan
Phone: 81-3-3812-2111 x5126
Fax: 81-3-3812-0544.
Location/Qualifiers
1. -5215
/organism="Pseudomonas chlororaphis"
/mol_type="genomic DNA"
/strain="B23"
/db_xref="taxon:333"
245. .248
/note="SD sequence for amidase"
254. .1774
/EC_number="3.5.1.4"
/codon_start=1
/transl_table=11
/product="amidase"
/protein_id="BAA14244.1"
/db_xref="GI:216851"
/translation="MAITRPTLDQVLDIRTLQHMQLTHEQAASYLELMQPSFDAYDVLV
LAGVPMNGAAPLEGVPGFDATVTVTRLLDAGATILGKATCEHYCLSGHSTSPAPV
HNPHRGVAGSSGSAALVAGSEVDIAVGGDGGSIPIPSAFCGTGKPKTEGLVP
YTGVAIETADIDHVPITGNVRDNLALQAMAGADGLDPRQAAPQVDYCYLKGVS
GLRVLGVEFALANQDPRVADKVRDAIARLEAGAHVPEVSIPEHNLAGLLMHPIC
EGLTQMNGHAGFNWGLXDVLDDQASWRDDAQLSASLKLCHMFVGGYLSRYN
GRYAKAQNLAARFQNGYDKALQTYLDDLMPTTITAOHPFANCSTIYVARALEMI
GNTAQDITIGHPAMSIPOCLLDGLPVLGMLVAKHYAEGTIYQAAAAFEASVDWRTL"
1857. .1861
/note="SD sequence for alpha subunit of nitrile hydratase"
1869. .2471
/EC_number="4.2.1.84"
/codon_start=1

FEATURES
source

RBS
CDS
Query Match 14.2%; Score 224.8; DB 1; Length 5215;
Best Local Similarity 48.8%; Pred. No. 3.9e-20;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;
Qy 37 CCACAGCCGACAGCTCCAGGAGTACAGCGCCGCCACCTTCGACTTCGACGAGGAA 96
Db 269 CCTACCTCCAGCAGGTTTAGACATCCGAAACCCAGTTCGACATGCACTGACGCGAA 328
Qy 97 CTGCGCCCGCAGCTCGCTCCGTCGTCGCGGAGATGGTACTGCTTCGACCTGATCGAC 156
Db 329 CAGGACGCTCTACTCTGAACTGATGCAACGAGTTTCGACGCCCTGACCTGTCGAC 388
Qy 157 GAACACTCCGCAACCGCGCGAGCGCGCTACACGCGACCGGACATCGGCCCGCGAA 216
Db 389 GAACGCTGATTTCGTTCCGCAATACGCTACGACCGCAGTTCAGGCTATCGCCATCG 448
Qy 217 CCACCGCGGACGACGCGCTTCACGCGATTCATCGGTTCTGCGCGGTGGAGCGCC 276
Db 449 CCATCGGCCAAGGAAACCCCTCTGAACGCGCTGTGTACTACCGAAGAGTGAATGTGCC 508

/transl_table=11
/product="nitrile hydratase alpha subunit"
/protein_id="BAA14245.1"
/db_xref="GI:216852"
/translation="MSTSIATSTPGRERAWALFOVLKXELIPEGVTEQLTQIMA
HWSPEAGARVAKAWVDQFRALLKKGATACAGQGYTGPGQETIYVALEDTPGVKNV
IYVSCJSCNFWPLGLPPEWYKGFREFRALVREGRTVRLBELTSPDVTIVKWDPTSA
ESRYLVLPORPEGSEHMBEQLQQLVTKDVLIGALPRVG"
2502. .2505
/note="SD sequence for beta subunit of nitrile hydratase"
2514. .3176
/EC_number="4.2.1.84"
/codon_start=1
/transl_table=11
/product="nitrile hydratase beta subunit"
/protein_id="BAA14246.1"
/db_xref="GI:216853"
/translation="MDGFHDLGFGQFGVPHPTINSLYKOVFKODWEHLAYSLMFVG
VPLKXPSVDEVRHAEVLDPQHYGTQYXVEYIATATLTAETGVIQAEIDQALGS
HPKLANPAHATGRPAITGRPPPEVGVVREYVAGHIRMAYVRGKEGVVHTTSE
QWPFDAIGHDLSAAHQPTYHVEPRVLDLWGDAAADDGIVVVDLFEYLKAPGAAV
NA"
3160. .3168
/note="SD sequence for P47K"
3173. .4432
/note="unnamed protein product; P47K"
/codon_start=1
/transl_table=11
/protein_id="BAA14247.1"
/db_xref="GI:216854"
/translation="MIEGAQAGRLPVTVLSGFLGAGTKTLNLAINRQGLRVAVIV
DNSEVNLDAESVQDVSLHRGDELTEMSNGICTRADLEQISDLAQORFDYLL
IESTGISPEMPVAETFAFDTRGSLSELRDLDTVTVVQSGFOALLSTDTVARAD
TEAHTSHLADLLEIQQEYANVLVNRKDLIDEPGYQAVHAILGALNPSASIMPAH
GNVALSLDHLFDLPASAQSGWNRKEDTFASESDDTGVTSWVTRKAPHPQ
RLLEFLQKPNHGRLLRSKGYFWLASRLHEIGLQAQSGKQFQWVGRWVNFIEPSQW
PRDEYLOGIMAKWDSVVGDCQELVFTGQGLDTRVLQRELDHCLLSAQEIAAGPLAW
QALPAATAPDTEALGARPTPPNAVQPT"
4439. .4443
/note="SD sequence for OrfE"
4450. .5215
/note="unnamed protein product; OrfE"
/codon_start=1
/transl_table=11
/protein_id="BAA14248.1"
/db_xref="GI:2160275"
/translation="MLFSTRTRWVLAPLADRLDOPDOACTPLSSNNPLRLRLTGE
QLLOEBSALKEQABALTOEVTOLDREVCEDRDLLEQWEARWALVSHGAGELFELEVN
GTRAPSLACAMRWTSASILAGIDQLGHNEQLPAPRQRHLDLTLARHLADRSRGP
FVLVDRIQPPSGDDYRWCRISGDSRRDAQGLPVMGGSURDIHQHQLQDEVLELATR
FDISREMLHDGLWDIEVWAGDPANPKNTIWWSSQMR"

QY 40 ACAGCGAAGAGTTCAGAGGTACAGCGCCCGCAGCACTTCGACTTCGACGAGGAACTG 99
Db 13334 ACGTGTTCGACACTGTGGGACATCTCTCGAACATCTTCGAGCGCTCGAAGAGTCCCT 13393
QY 100 GCGCCGAGTCTGTTCCGCTCGTGGGAGATGTTGACTGCTTCGACCTGATCGACGAA 159
Db 13394 GAAGTCGACCGGAGCGGAGCTGTGTGACGTGATCGCGGCCGACGAGATCCGACTCC 13453
QY 160 CTACCGCAACCGCCGAGCGCGCGACGCGGTATACGGAACGGGACATCGCGCGGAAACGG 219
Db 13454 CTCGACGAGACGACGCGCTCGCAACGCGCCCGCAACCGGAGCGCGGTTCGAAAGCG 13513
QY 220 ACCGGG--ACGAAGACCCGTTCAACGCACTTCATCCGTTCTCGCGGTGGAGGGCGCA 277
Db 13514 CCCAAGTCTGTAAGAGCTGAACGCACTTATACCCGAGAGCACTTCGAGGAGCGG 13573
QY 278 CGAGGGGCACTGAGCGCACTGACCGCGCGCATCAAGGACTGCATCGCCATCGCGGTA 337
Db 13574 CGACGCGCCCTTCGACGCGCACGACGCTGCGCTCAAGGACCAATCTCCAGAAAGCGG 13633
QY 338 TGCCCAACGACGAGCGGTCCGAGTCTCCGACTCTCGACTGTGATCGCCACCGAGGATCGCGTGG 397
Db 13634 TGCCCAACGAGTGTGCTCGAAGTCTGCGCGACTACGAGCGCGCTGATCAACGCCACCG 13693
QY 398 TGGTGGAGCGGTGCTCGCGGACGGCGCCACCATCGTTCGCAAGAGCAACCTCGAGGACA 457
Db 13694 TCGTCAGGACCTCTCTCGCGCGCGCGGACCATCGTTCGCGCAAGACCAACATGAGCGAGT 13753
QY 458 TGGCGATGGTATCGGTGAAGGACGCTCTA---CGGTCTCGCTGTAACCCGCAACAC 514
Db 13754 TCGGATGGGAGCGACGACGACCTCGTACTTCGCTCCGACGAAGACCCGGTGCAG 13813
QY 515 CGCCCAACGACGAGCGGTGATCTTCAGAGGGCTTCGCGCGCTGCGCTGCGCGGATCG 574
Db 13814 AACCCTGGTTCGCGCGGCTCTCTCGCGGCTCCGCGCGCGCTCGGACGCGGACG 13873
QY 575 TCAGTTCGCTTGGGCTCGATGAGGAGGAGCATTCGATTCGCGCGCGATGTTGCG 634
Db 13874 CGACCTCGGCTCGCTCGACACCGCGGCTCGGTTTCGCGCGCGCGCGCTACTGTG 13933
QY 635 GACTGTGCGCATGAAGGACCCAGCGCTGCTGCTTACGCGCTTACATACATCG 694
Db 13934 GCGTGTGCGCTCAACGACGATGAGGCTGCTTCCGCTGACGACTGTTGGGTAG 13993
QY 695 ACCACACTTGGACCAATCGCGCCCATACACGAGGGGTCGAGCTCAACGCGCGGTC 754
Db 13994 CAACAGCTCGAAGAGATCGGCGCGCTCGCGCCACCGCTCGAGGCGCGGCACTCC 14053
QY 755 TCAGGTTGTCGCGCGCGGCGCATGCGCGACCTCAGTGGTGCCTAACCTTCGAGAC 814
Db 14054 TGACGTGATCGCGCGCGCGCGACGACGACGCGGACCCCGGACGCGCGGCGGAGCG 14113
QY 815 CGGAGAACTACGCTTCGCGCTCGCGAGGAGGATATCCGCTCGAGATTTCGCGTCTCG 874
Db 14114 CTGACTACGCGAGCGCGCGCGCGCA-----CGTCGATGGTTGACCATCGCG 14164
QY 875 AGAGTCACTGGAGCGCAAGGTCGACGCGCGGACGATGATCGCGGTTCAACAGGAC 934
Db 14165 TCCGCGAGAACTACCGAGCGCGCTGACGACCGCTCGTTCGAGCGCTTCGAAGCGCGC 14224
QY 935 TGGCGCGCTCGAGAGCGCGCGTCCGACCATCGAGCGGCTCTCGTCCGCTTGTGACGG 994
Db 14225 TTGCGGACCTCCGCGGAGGCGCGACACCTCGACGTGGGTTGCGCTCTGTGAT 14284
QY 995 CGGCTGGCTATTCGAGAGGCGGTGATGGCTTCAACGCGCGCGCTTGGCGGATCGCG 1054
Db 14285 ACSCGCTCGCGCTACTACGCTATCGCCATGTCGCAAGCTCTCGAACCTCGCGCT 14344
QY 1055 CCGTGTGGGCTACTTCCACAAGGGCGCGTGGAGCTCAGACCGCGCTCACACGCGCG 1114
Db 14345 TCGAGCGGCTCGGTACGCGACCGCGCGCGCTTGAACGCACTGGAACGAGACGCTTCG 14404
QY 1115 CCCAGAGTCGACCAACCCACAAGGATTCGCGGATCTGTCGCCGCTGATGCTGTTGATCG 1174

Db 14405 CGGACGCGCGCGCGAGCGGGTTGGGACGAGGTTCAGCGCGCATCTCTCTGGTA--- 14461
QY 1175 CGGAGCACTTCGCGCGAGCAATACCTCGGCATCCACTACGCGAAGCGCGACAGACCTCGCGC 1234
Db 14462 CGTACGCGCTGTGCGCGGCTACCGACCAATACTACAAGCAAGCCAGAGCCCGGG 14521
QY 1235 TGGAGCTCGGCAAGCAGATCGACCGCTCTCCAGGACCGGGTGCATCTGTCGACCCGGA 1294
Db 14522 CGTGGGTGAACAGGACTTCGACGAGACGTTTCGCGGACGTCGACGTGGTGGCGTCCCGGA 14581
QY 1295 CCAGCTTACGTTGCGCAAGAGCTGTTAGCGGTTCGCGAAGCACCACATGTCATGATCC 1354
Db 14582 CGATGCGGTCTCTCCCGCGAGGCTCGGCGAGGCTCGAGACCCCGTCCAGATGATC 14641
QY 1355 CACGATGACCGGCAATTCGATCTCAACAGTTCGCGCTGAGCTTCGACCTCACGGTCAACCGG 1414
Db 14642 -----TCGCGGACGCCAACCAACCGCGCGGCAACCTCGGAAACCTCCCGC 14686
QY 1415 CGCTGACGCTGCCACCGGTCGCGGCGAGAGGCGCTCGCGCTTGGGCTCCAGTGATAG 1474
Db 14687 CCATCTGCTGCGCGCGGACGCG---ACGGCTCCCGGTGCGGCTTCAGCTTGTGCG 14743
QY 1475 GCCGCCACTTCGAGAGTCGACGCTTACCGCACCGCGCGCGTATCGA 1523
Db 14744 GCCGAAAGTTCGCGGAGGAAACGATCATCAACGCGCGCGCGCGTCTGA 14792

RESULT 8
SC0939124 308050 bp DNA linear BCT 11-FEB-2003
LOCUS Streptomyces coelicolor A3(2) complete genome; segment 21/29.
DEFINITION AL039124 AL031031 AL031035 AL031124 AL031225 AL031231
ACCESSION AL031260 AL034447 AL035559 AL035569 AL0355913 AL0355913 AL0355913
VERSION AL039124.1 GI:2413886
KEYWORDS Streptomyces coelicolor A3(2)
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 Bentley, S.D., Chater, K.F., Cerdano-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neill, S., Rabinowitch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J. and Hopwood, D.A.
Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)
JOURNAL Nature 417 (6885), 141-147 (2002)
MEDLINE 21996410
PUBMED 12000953
REFERENCE 2 (bases 1 to 308050)
Bentley, S.D.
Direct Submission
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
On or before Oct 26, 2002 This sequence version replaced gi:20520664, gi:20520751, gi:20520752, gi:20520665, gi:20520811, gi:20520756, gi:20520845, gi:20520760, gi:20520817, gi:20520818, gi:20520782.
FEATURES Location/Qualifiers
1..308050
/organism="Streptomyces coelicolor A3(2)"
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
250..2490
/gene="SCO5495"


```

CDS
/note="synonym: SC8D9.07"
250. .2490
/gene="SC05495"
/note="SC8D9.07, possible membrane associated phosphodiesterase, len: 746 aa; similar to many eg. TR:O87376 (EMBL:AF052518) c-di-GMP phosphodiesterase A from Acetobacter xylinus cdg2 operon (752 aa) fasta scores: opt: 959, z-score: 1059.2, E(): 0, (31.7% identity in 603 aa overlap). Appears to have two distinct regions with the first 300aa containing several potential membrane spanning hydrophobic regions and the remainder of the protein showing similarity to the phosphodiesterase mentioned above. Also similar to a nearby gene on the same cosmid (SC8D9.22) fasta scores: opt: 736, z-score: 625.6, E(): 5.3e-30, (32.6% identity in 729 aa overlap). Contains TTA Leu codon, possible target for bida regulation. Also contains Pfam match to entry PF00390 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46 and Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83. Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative phosphodiesterase"
/protein_id="CAB37571.1"
/db_xref="GI:4467246"
/db_xref="GCA:Q92584"
/db_xref="SPTREMBL:Q92584"
/translation="MEPTESAAAGSLRLRRVADAMHVSRLGQDGVFSSMTGTRV PCAYGRPHGPHGLVDPDGRHLSWPALPAVATPAFVLGAGFYAFSDGHAL FPAGTVGSLAVLSGIVVGLHVLMLGRSRMGWGSGAALTVALLXGWTGAGVVSUT VVVLVGIAARRGRQVGLHVGADILGIAAGAVLGVCGVSPVERPMDPETHVYAAP KVLIVAVAYLAVTALLWYITOTRPGPLTAARTALVRHGLVAVALLGIAPLCVAVA KELLIPFAIPLIALDSLTWTARASEQLRDLPLTGNKOLLERTWTALDDERIG ARAAMLIDRFVSNLDLHAGDRLLQTLADRLKQALPRGAERAEQDADFAVLL PVADSTIARSLVSGALSPMDLGLTLVLEASGAVVFPFDALDAGLLRRADV AYAKQDRDTGVVEVSRDNTPDRLGLDLRRALDAHEVOLHYQFVRFQDQVAG LEALVVRHPRGVKPPDFETAIASEGLMPLTEYVLEALGVQAWRAQGLFVPVA VNSPRDVHTFGFAGSVAARLGRVFPAGALQITEHVLEDPQARADLTNALTGHG VKMSLDGFGTSSLVHLRLPVSELKTDRSFVARLADNEDAEIVRCTVDLAHSLGL LVAAEGVEDDETWERLRLDGCDAVQGVLMVAAMPPEETATALLARGSGVRAAAALP AASDDE"
1141. .1635
/misc_feature
/gene="SC05495"
/note="Pfam match to entry PF00990 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46"
1420. .1422
/misc_feature
/gene="SC05495"
/note="TTA Leu codon, possible target for bida regulation"
1672. 2406
/misc_feature
/gene="SC05495"
/note="Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83"
2774. .3145
/misc_feature
/gene="SC05496"
/note="synonym: SC8D9.08"
2774. .3145
/misc_feature
/gene="SC05496"
/note="SC8D9.08, small hydrophobic membrane protein, len: 123 aa; no significant database similarities. Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative small membrane hydrophobic protein"
/protein_id="CAB37572.1"
/db_xref="GI:4467247"
/db_xref="SPTREMBL:Q92583"
/translation="MTVTATYLLAVVNLNFCFLGLGRFLFPQASATGVGVDPADGGDA AAYLVKGVDRDGTGLGVGLLAVLAFAGARPEAFMCLVALVPLADTLVLRHGGTKAVA FGIHFAITAVLVLSAGLLFAV"
3177. .3671
/misc_feature
/gene="SC05497"
/note="synonym: SC8D9.09"
3177. .3671
/misc_feature
/gene="SC05497"
/note="SC8D9.09, unknown, len: 164aa; similar to many of undefined function eg. TR:O06336 (EMBL:Z95390) hydrophobic protein from Mycobacterium tuberculosis (177 aa) fasta scores: opt: 171, z-score: 213.4, E(): 0.00014, (34.9% identity in 109 aa overlap)."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAB37573.1"
/db_xref="GI:4467248"
/db_xref="SPTREMBL:Q92582"
/translation="MSLFPVKDESVVVREAEVVGRAFTVRLIADSSSGGALST QVTVLTAGDAGKAPKPHDMSAMFFLLDGAADVLSDEVLTAGPGGLIVVPPKPHAF AAVPGADALLVLAPGVFERFEYRHLQRIALGEATPESLLEVQELYDNHFLSSAAK AARG"
3816. .4112
/misc_feature
/gene="SC05498"
/note="synonyms: gatC, SC8D9.10"
3816. .4112
/misc_feature
/gene="SC05498"
/note="SC8D9.10, probable Glu-tRNA-Gln amidotransferase subunit C, gatC, len: 98aa; similar to many eg. TR:O06492 (EMBL:AF008553) Glu-tRNA-Gln amidotransferase subunit C, gatC, from Bacillus subtilis (101 aa) fasta scores: opt: 232, z-score: 294.8, E(): 4.1e-09, (42.1% identity in 95 aa overlap)."
/codon_start=1
/transl_table=11
/product="probable Glu-tRNA-Gln amidotransferase subunit C"
/protein_id="CAB37574.1"
/db_xref="GI:4467249"
/translation="MEGITREEVAHLARLEKPEELEHFAAGOLDLIGAVARVSE VADQVPFTSHPLPLTNVWRPDEVPSLTPEQALSGAPAQEQRRFKVQILGEE"
4116. .5609
/misc_feature
/gene="SC05499"
/note="synonyms: gatA, SC8D9.11"
4116. .5609
/misc_feature
/gene="SC05499"
/note="SC8D9.11, probable Glu-tRNA-Gln amidotransferase subunit A, gatA, len: 497aa; similar to many eg. TR:O06491 (EMBL:AF008553) Glu-tRNA-Gln amidotransferase subunit A, gatA, from Bacillus subtilis (486 aa) fasta scores: opt: 1443, z-score: 1512.5, E(): 0, (52.0% identity in 465 aa overlap)."
/codon_start=1
/transl_table=11
/product="probable Glu-tRNA Gln amidotransferase subunit"
/protein_id="CAB37575.1"
/db_xref="GI:4467250"
/translation="MSDIKLTAEAKIAGELTAQVTEHLARIENAVDEKVAHF LHYDRGALQAQAVDEKREKGLAGVPLAKDIFTEGIPTVGSKILEGWIP PYDATVTKLADVVILGKTNNDPEFANGSTENSAYGTGNPMDLTKIPGSGGGG AALAAFAQILAIGTDGTGSTRQPSVTVGVKPTVGVSGVAFSSLDQGGPCA RTVLDALLHEVIAGHPMDSTSIDAPVAVVAARNGSVGDMRNVGVQFRGEGYA GVQVRDESVELLKSGIAEVELDCPSFDLALSAYTLIAPSECSNLAEPDGLRYGAR VGDGTHSAEVTSLTREAAGFQGVKERRIMLGTALSSGYDAYGSAOKVRLTIQKE FERAFQDVIVISPTPTTFAIGERADPDPMANVLDLCTIPTNLGNAAMSIPCLGLA PENLPLVGVQITAPAKMDRLKYGAAVAAAFVEKKGHPHLEIAPSL"
5606. .5845
/misc_feature
/gene="SC05500"
/note="synonym: SC8D9.12"
5606. .5845
/misc_feature
/gene="SC05500"
/note="SC8D9.12, putative membrane protein, len: 79 aa. Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB37576.1"
/db_xref="GI:4467251"
/db_xref="SPTREMBL:Q92579"

```


COMMENT

This work was done in collaboration with Haruo Ikeda(*1), Jun Shikawa(*2), Akiharu Hanamoto(*3), Chikazu Takahashi(*3), Mayumi Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi Osone(*4), Norihiro Kuehida(*4), Hisashi Kikuchi(*4), Tadayoshi Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7) and Satoshi Omura(*1,*3).

Final finishing process and all annotation were done by H. Ikeda and J. Shikawa.

- *1 Kitasato Institute for Life Sciences, Kitasato University
- *2 National Institute of Infectious Diseases
- *3 The Kitasato Institute
- *4 National Institute of Technology and Evaluation
- *5 School of Science, Kitasato University
- *6 Institute of Medical Science, University of Tokyo
- *7 RIKEN, Genomic Sciences Center

Following url is also available.

<http://avermiltilis.ls.kitasato-u.ac.jp>.

FEATURES

source

1. 281450

/organism="Streptomyces avermitilis MA-4680"

/mol_type="genomic DNA"

/strain="MA-4680"

/db_xref="taxon:227882"

/note="This strain is also named as strain: ATCC 31267,"

NCIMB 12804 or NRRL 8165.-synonym: Streptomyces

avermectinus"

359. .4195

/gene="cvaA6"

359. .4195

/gene="cvaA6"

/gene="SAV2694"

multi-component regulatory system-6"

/codon_start=1

/transl_table=11

/product="putative sensor-like histidine kinase"

/protein_id="BAC70405.1"

/db_xref="GI:29606343"

gene

CDS

```

RADQAAVAVASGLTSLTAGASRIFFEGGTVAQTVMEMERGFLFLMSVSDGSLAVLSHPE
CDIGLVGEMALLVDRAGAVLTPDLRAELQGSLLH"
4730. .5128
/gene="cvaC6"
4730. .5128
/gene="cvaC6"
/note="SAV2696
multi-component regulatory system-6"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC70407.1"
/db_xref="GI:29606345"
/transl_table=11
L1STTADPAQLMGLLPEHQRIChLCEVKSVAEVSALLSMPGLVARILVADLAAGLV
AIHQPGDENNGGAPDVTLLERVLGLRLK"
5109. .5687
/gene="cvaD6"
5109. .5687
/gene="cvaD6"
/note="SAV2697
multi-component regulatory system-6"
/codon_start=1
/transl_table=11
/product="putative ATP/GTP-binding protein"
/protein_id="BAC70408.1"
/db_xref="GI:29606346"
/transl_table=11
AVMTSASAGIDLLTHGDKTTTVMDFGRITLDQDLILYLFGTQDGRFWMFDLLV
RGAIGAVLVDRRLADCPFVDPYFENSGLPFVIALNGFDGHPQYNPEEVRALQIGP
DAPITTDARHADRAKASLITLVEHALMARL"
6085. .9330
/gene="cvaA7"
6085. .9330
/gene="cvaA7"
/note="SAV2698
multi-component regulatory system-7"
/codon_start=1
/transl_table=11
/product="putative sensor-like histidine kinase"
/protein_id="BAC70409.1"
/db_xref="GI:29606347"
/transl_table=11
PRNRYVRLNALILLPVLVLMGFGVQKSGISIDTQWEARDAEKVAKIVASAGEYABA
LIMERVVSAPALLKGRNNATVKKAFALTQKADAFPHQVGMGPAGLORLGLVEK
AEPQLQIKAAVYTTAMDVPKTEGVTVEHLIMEFSENLGLTGNITSYGRITVAIA
LAGASLRSIRIGLHLLIKPSSDEKIRAQOITSFTSYALEQIALQEVYSGGTEODYA
ELTQWQKQTAAGKAKLAKSRTAAGTFAVAPALDKMIOAIGSATPAALAOQGI
RTPWMSAAILKFDGYSQVEKELVDVANDANDIASSAQDAYINGAIVVVALAIFI
LAGLMARQSRMRQURNAFVNAEORLEMLVDQLSRDTPGRVDTIRVQAIPIITTDEI
GEVAPADQVHREAVLAAEQALLRGNINAIPTNLSRRNSLIEGQTLITDLENNEA
DPQENLFLDLHATMRNGENLLVLAEEFGRWDQFVLVDVLRASSEVEQYI
RVLSGVPAEAIHGRAVITDLVHLLAELENATTFSSPQTKRVTRATRLPDGRIMVRIH
DKIGLITADPADINHLANPPTVDAAIISQRMGLFVVGRLSDRHGIRVLRPSGEQAG
TTSVLVLPDVIITHGGGGEQVORDEFTSQIIPQONFAGGGEAFQVPMRTAAELGF
DDREYTEVDDTRELDPVRSLSNREERAALEAHPQDSEGGQYQDEASQCPA
YGMQNGYEQQAGYQQQYAYEQQQPAIZEEYFANGFSQTEGEPAGEFAPNGDP
ANGGVPESYAPFENSGSGRTASFTSFAPDFSYQDDMPKQDDISGYSKSEVAP
EAESSQAADYSERDHFVFTPGPAPSAHALTDAGLPRRGSTNGNGNGQVRVPPDS
AAPAGTETGMSGNRSNDRSQASQLKPKAGGVTSSGLPRRYVKANLVGTADS
TPQGQPVQSRPADVRGRLSNLRGVRGRNAGSDTNGQLGPDSTYNQER"
9341. .9754
/gene="cvaB7"
9341. .9754
/gene="cvaB7"
/note="SAV2699
multi-component regulatory system-7, containing
roadblock/LC7 domain"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC70410.1"
/db_xref="GI:29606348"

```

gene

CDS

```

/translations="MSQAAQNLMNLTNFDVNTPGVSHTVVVSADGLLILAMSEGGPRD
RADQLAAVASGLTSLTAGASRIPEGGAVNQTVVEMERGFLEFLMSISDGLSLAVLHPE
ADILGLVGYEMALLVDRAGTIVTPDLRAELQGSLLN"
9851..10435
/genes="cvmC7"
CDS
9851..10435
/genes="cvmC7"
/notes="SAV2700
multi-component regulatory system-7"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC70411.1"
/db_xref="GI:29606349"
/translations="MATPPGGSSGNWSYGPQGQGGQBNRYNFPSPAPRRQPYPAP
Query Match 13.2%; Score 208; DB 1; Length 281450;
Best Local Similarity 50.1%; Pred. No. 1.8e-18;
Matches 641; Conservative 0; Mismatches 615; Indels 24; Gaps 4;
QY 262 CGGTTGAGGGCGCCACGAGGGGCCACTGAGGACCTGACCGCGCGATGCAAGGACTGC 321
Db 72139 CGGAGCGCGCGAGAGCTCGGCCGCTCGCGCGCTCGCGCTCAAGGACATC 72080
QY 322 ATGCCCATCGCGTATGCCACACGAGCGGTCCCGGATGCTCCGACTGTGATCGCC 381
Db 72079 TTACACACGAGGACATCCGACACCGCTCGGTTCGAAGATCTCGAGGCTGATCCCG 72020
QY 382 ACCGAGGATCGGTGTGGAGCGGCTGCTCGCGGCGCGCCACCATCTGTCGGCAAG 441
Db 72019 CCTACGAGCGGACCTCACCAGAAAGCTGAAGCGCCGCGACGTGCTCATCTCTCGCAAG 71960
QY 442 ACAGACCTCGAGGACATGCGGATGG---TATCGGTGAAGGACGCTACGCTCTCTGG 498
Db 71959 ACCAACATGAGGAGTTCGCGATGGGTCTTCACCGAGAACAGCGGCTACGCGCCAC 71900
QY 499 CTGAACCGGAACACCCCGCCACCGCAGCGGTGGATCTTCCAGCGGCTCCGGCGTGGC 558
Db 71899 GGCAACCCCTGGGACCTCAGCGGTATCCCGGTGGCTCCGCGGTGGCTCTCCGCGGC 71840
QY 559 GTCGTGGCGCATGTGTGACTTTCGCTGCGGTGGGTGATGAGGAGGACGACGTCGATC 618
Db 71839 CTCGCTCTTCGAGGCGCCCTCCGCTCGGACCGGACCGCGCGCTCCATCCGCGAG 71780
QY 619 CCGCGCGCATGTGCGGACTGTGCGATGAAGCGGACCGACCGCGCTGTGGCGCTTTC 678
Db 71779 CCGCGCGCGCTCAGCGGACGCTGCGGTGAGCGGACGCTACGCGCGGTCTCCGCTAC 71720
QY 679 GCGCTGACATACATGAGACCAACCTTGACACATCGGCGCCCATCACAGGGGGGTGAG 738
Db 71719 GGCATGTGTGGCGTTCCTGCTCTCCCTCGACCGGCGCGCCCTGTGCGCGTACGCTCTG 71660
QY 739 CTCAACGCCGGGCTCTCGAGGTGTGCGGGGCGCGACTGGCGGACCGCTCAGTGGGTG 798
Db 71659 GACGGGCGCTGTGACAGAGTATCGCGGGGACAGACCGGCTCGACTGAGTCCAT- 71601
QY 799 CGTAACCTTCGCGAGCGCGAGAACTACGGCTTCGCGTTCGCGGAGGAGTATCGGCTG 858
Db 71600 --CGAGCGCGCGTCCCGCGGTGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGATG 71543
QY 859 AGATTGCGGTCGTCGAGGAGTCACTGAGCGGACCGTGCAGCGCGGATGATCGCC 918
Db 71542 CGCGTCGCGTCTGCAAGAGTTCGCGCGGAGGGCTACACAGCGCGGTGTGCTCCAGCG 71483
QY 919 GCGTTCAACAGGAGTGGCGCGCTCGAGAGCGCGCGTGGACCATCGAGCGGCTCTCG 978
Db 71482 TTCGAGAGTCCGTCGCGCTCTCAAGAGGTGGCGCGGAGATCGTTCGAGTGGACTGC 71423
QY 979 GTGCGGTGTGAGCGCGCGCTGCGCTATCCAGAGCGCGGTGATGCTTTCAGCGCGCG 1038
Db 71422 CCGTCTTCGACCTCGCGCTGTGCGGCTGCTGAGTTCGCGCGCTCGAGTGTCTCTCC 71363
QY 1039 GCTATGCGGAGTTCGCGCGGTGTGGGTACTTCCACAGGGGCGCGTGGAGCTCAGCAC 1098

```

```

71362 AACCTCGCGCTTCGACGGGCTCGCTATGCGCTGCGGTGGGTGACGACGCGACGCG 71303
QY 1099 GCGCTCACGCGCGCGCCAGAGTCGACACACACACACAGGATCTGGGATCTCTGCGG 1158
Db 71302 TCCGCGGAGGAGGTACCGCCCTCACCGCGAGCGGGCTTCGCGGAGGAGTCAAGCGC 71243
QY 1159 CTGATGCTGTGTGATCGCGGAGCACCCTCGCGACGAATACCTTCGCGCATCTCAGTAC 1218
Db 71242 CGCATCATGCTCGGACGCTAGCTCGGCTCGCTACTACGACGCGTACTACGCGGACG 71183
QY 1219 GCGGAGAACTGCGGCTGGAGCTCGGAGAGCAGATCGACGCGCTCTCCAGGACGCGG 1278
Db 71182 GCGGAGAACTGCGGCTGGAGCTCATCACCGCGACTTCGAGAAAGTCTGTCGACGAGT 71123
QY 1279 GCATGCTGACCGCGGACGACGCTTACCGTTCCCAACGAGCTGTTGAGCGGTGCGCAAG 1338
Db 71122 GTGATCGTCTCCCGGACGACGCGCCACACCGCTTCCCGATCGGCGAGCGCGGACG 71063
QY 1339 ACCATGTCATGATCCGACGATGACGCGGATGACGCGGATCGGATCTCAACAGTGCCTG 1398
Db 71062 CGATGCGGATGTACC-----TCGCGGACCTGTGCACCATCCCGACCAAC 71018
QY 1399 CTCACCGGTCAACCGCGGCTGACGCTGCCACCGGT---GCGGGCGAGAGCGCTGCC 1455
Db 71017 CTCGCGGCAACGCGCGCATGTGCTGCTCGCTCGCGCTCGCGCGGAGACAACTCCG 70958
QY 1456 GTTGGCTTCAAGTATAGGCGGCACTTCGAGGAGTGCAGCTCTACGCGACGCGCGCC 1515
Db 70957 GTGCGCTCGCATCATCGCGCGGCTCAAGGACGACAGGCTGTACAAGTCTCGCGCC 70898
QY 1516 GTGATCGAGGCGCGCGCT 1535
Db 70897 GCGTTCGAGCGCGCTTCTGT 70878

RESULT 10
LOCUS AE016864
DEFINITION Pseudomonas syringae pv. tomato str. DC3000 section 9 of 21 of the
complete genome
ACCESSION AE016864
VERSION AE016864.1 GI:28852711
KEYWORDS
SOURCE Pseudomonas syringae pv. tomato str. DC3000
ORGANISM Pseudomonas syringae pv. tomato str. DC3000
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 310325)
Buell, C.R., Joardar, V., Lindeberg, M., Selengut, J., Paulsen, I.T.,
Gwinn, M.L., Dodson, R.J., DeBoy, R.T., Durkin, A.S., Kolonay, J.F.,
Madupu, R., Dougherty, S., Brinkac, L., Beanan, M.J., Haft, D.H.,
Nelson, W.C., Daviden, T., Zafar, N., Zhou, L., Liu, J., Yuan, Q.,
Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K.,
Utterback, T., Van Aken, S.E., Feldblyum, T.V., D'Ascenzo, M.,
Deng, W.L., Ramos, A.R., Alfano, J.R., Cartinhour, S., Chatterjee, A.K.,
Delaney, T.P., Lazarowitz, S.G., Martin, G.B., Schneider, D.J.,
Tang, X., Bender, C.L., White, O., Fraser, C.M. and Collier, A.
The complete genome sequence of the Arabidopsis and tomato pathogen
Pseudomonas syringae pv. tomato DC3000
Proc. Natl. Acad. Sci. U.S.A. 100 (18), 10181-10186 (2003)
12928499
JOURNAL 2 (bases 1 to 310325)
PUBMED Buell, R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D.,
BERRY, K., Utterback, T., Van Aken, S., Feldblyum, T., Gwinn, M.,
DODSON, R., DeBoy, R., Durkin, A., Kolonay, J., Madupu, R.,
DOUGHERTY, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J.,
NELSON, W., Daviden, T., White, O., Fraser, C. and Collier, A.
Direct Submission
JOURNAL Submitted (03-MAR-2003) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
Location/Qualifiers
source 1..310325

```


GVFAPSGKHLRTVPAGERIITTAGELIKVESIKOIAOFPSNEKGTLSIATHTT
GARYALPVSINFIKQYEDVALHMQSPMOIAEWAADGTVDPAIATBALEQRDLWM
MCIYRNKCVVFPQGRHLTKVGLKLELLASIPVIVYFGTIGKGLDEAFSHRGLVP
KVFTAADADIKTKTVRLNLGVIGVARVAVDEVDSDLVLDLDFESSVTIKAFRR
GTFRLGFMDFIKTFAPHLTRVWAKA:QCHNKQEMSELFANVELPVH"
complement (7787. .8521)
/gene="cysh"
/locus_tag="PSPTO2280"
complement (7787. .8521)
/gene="cysh"
/locus_tag="PSPTO2280"
/codon_start=1
/transl_table=11
/product="phosphoadenosine phosphosulfate reductase"
/protein_id="AA055794.1"
/db_xref="GI:28852722"
/translation="MSQPPDVAELATTYANKSADILKLAHEHFGDELWISFGAEDV
VLVDWANKLNKVKVPSLDTRLHPETVRFIEQVDRDHYKIDIELISPPQRALEPFVKE
KGLFSYRDHGECGCKIEPLRLKLSVKAATGQRRDQSPGTRSAVALEVDSEAF
Query Match 13.0%; Score 205.8; DB 1; Length 310325;
Best Local Similarity 50.8%; Pred. No. 3.4e-18;
Matches 632; Conservative 0; Mismatches 592; Indels 21; Gaps 5;
QY 210 CCGCGAACCGACCGGGCGAGACACCGCTTCAACGATTCATCCGTTCTGCCGGTGA 269
Db 156648 CAGCAAGCCGGCAGATGCGCAAAACCTGCACGGTGCCTGGTACGTAAACCAACATCGT 156707
QY 270 GGGCGCACCGAGGGGCCACTGAGCACCTGACCGCGCGGATCAAGAGCTGCATCGCCAT 329
Db 156708 CGGTGGCCAAACGGTCTCTGGCGGAGAACGGTGGTTCATCAGGACAAAGTTTCAGT 156767
QY 330 CGCGGTATGCCACCAAGAAAGGGTCCCGGATGTCCTCCGATGTGATCGCCACCGAGGA 389
Db 156768 GCGCGCGCTGCCGATGGCTAATGGTTCTTTGAGCCTTGATGGCTACGTGCCGCTCGAAGA 156827
QY 390 TGCCTGTGTGTGGACGGCTGCTCGCGCAGCGCCACCATCTCGTCCGCAAGACGAACCT 449
Db 156828 CGCACGGTGTCAACAGCCTGTTCGCGCTGTGTGCGAGCGGTGCTCGGCAATCATGTGTG 156887
QY 450 CGAGGACATGGCGAT---GGGTATCGGTGAAGCAGCGCTTACGGTCTGCGCTGAACCC 506
Db 156888 CGAAGACCTGTGTTCTCGGGAGCCAGCTTTACCTCGGCGAGCGCGCGGTGAAGAAACC 156947
QY 507 GAACAAACCGCCGACGGCAGCGGTGATCTTCCAGCGGCTCGGCGCTCGCTCGCTGC 566
Db 156948 GTGGGACCTGGCCCGCAACCGGGCGGCTCGTCCAGGGGAGTGCCTGCTGCTGCTTCGT 157007
QY 567 CGGCATGTGCTGACTTCGCCCTTGGGCGTTCGATGAGGAGGAGCAGCATCCGGATCCCGGCCG 626
Db 157008 GCAAGAGGTGCATGSCCATTTGGCGCGATCAGGCGGTTTCATCCGTATGCGGTCGGC 157067
QY 627 ATGCTCGGACTGTTCGGCATGAGGGACCCACGGCCTGTGCTGCTTACGGCTGAC 686
Db 157068 GTGAGCGGCGATCGTGGGCGACAGCGGAGCTGACGCTGTGCTGCTTACCGGCTGCTTT 157127
QY 687 ATACATGACACACCTTGACACACATCGGGGCCCATCCACAGGGGGGTGAGTCAACGC 746
Db 157128 CCGGATTGAAGCACCACATGACCATGTGCGCCGATGGCCAAACAGTGGCGACTGCGG 157187
QY 747 CCGGCTCCTGAGGTGTGGCGGGCGGACTGGCGGACCCCTCAGTGGTGGTGAACCT 806
Db 157188 GGTGATGCTGACGTGATTCGGGTGCTGACGCTCTCGAC-----TCAGGCGAGAAGAA 157241
QY 807 TCCGAGCCGAGAACTACCGGCTCCGGGCTCGGCGAGGAGTATCCGCTCTGAGATTCG- 865
Db 157242 TCCGCCAGCGGTGAGTTGTGTGCGGACCTTGATCAAGGGTGTGCGGCTGAAATCGG 157301
QY 866 --CGGTGCTGAGAGGTCACTGGAGCCGAACGGTGGACCCCGGACGTGATCCCGCGTT 923
Db 157302 TTTGTTGGCGCAAGGTTTCGCCATCCCGGCATGTCGAGCCGCGAGTGGATCGCTGCT 157361
QY 924 CAACCGGAGCTGCGGGGCTCGAGAGCGCGGTGCGACCATCGAGCGGCTCTCGGTGCC 983

Db 157362 CAAGCTGCGGTGGCACAACTGCAAGCCTCGGGGCCACGGTGGCGAGGCAAGCGTGC 157421
QY 984 GTTGTGAGACCGCGCCTGGCTATCAGAGCGCGGTGATGGCTTCAACCGCGCGCTAT 1043
Db 157422 GTGCGCACTGTTGTGGCGCTGGACATGTGGAAACGTATCGCCACTGACCGCGCGCGTA 157481
QY 1044 GGGGAGCTCCCGCGGTGTGGGCTTACTTCCACAAAGGGCGCGGTGGAGAGTCAAGCACCGCGCT 1103
Db 157482 CCAGATGCTCAAGGCAATGGCTACGGCATGAAGTCAAGCGGCTACTACGACCCAGAGAT 157541
QY 1104 CACGA-----CGCGGGCCAGAGTGCACACCCACAGAGATCTGCGGATCTCTGTC 1157
Db 157542 CATGAGCTACTTTCGGTCCCAAGCGTCCGCAACACGCAACACGCTTGTCCAGCAGCGTACG 157601
QY 1158 GCTGATCTGCTGATCCCGGAGCACCTGCGCGACGAATACCTCGGATCCAGTATCCATACGCGAA 1217
Db 157602 CGTGTGGCGCTGACCGGCTATTACAGCTGAACACCTGACAGGTCGCTTACTACGCCAA 157661
QY 1218 GGCACGAACCTCGGGTGGAGTTCGGCAAGCAGATGACGCGCTCTTCAGGACCGGGC 1277
Db 157662 GCGCGCATGCTGTGTCGGAACTCACTCGCCAAATACGACGAAGCGTTCAAGGACTTCGA 157721
QY 1278 TGCATCTGACCCCGACCGCTACCTGCTTGCACAGAGCTGTGAGCGTGTGGAGCGTTCGGCAAGA 1337
Db 157722 CGTCTGCTGATGTCGACCATGCGCTTGTGCGGAGC---CCGTTGACCGCGCGATGC 157778
QY 1338 CACCATGTCCATGATCCACGATGACGGCAATGCGATCCTCAACAGTGCCTGGA 1397
Db 157779 GCCGATTGAAGAGTACGTGACAGCGCGCTGAACATCTCGCAACACCGCACCGTTCGA 157838
QY 1398 CCTCACCGGTACCGCGGTGACGGTGCACCGGTGCGGCGA 1442
Db 157839 CCTGACCGGCCACCGCGCACCTCGATCCCGCGGCTCTCGCGA 157883

RESULT 11

REAG Rhodococcus rhodochrous amidase gene.
LOCUS Rhodococcus rhodochrous 1958 bp DNA linear BCT 04-FEB-1999
DEFINITION D16207
ACCESSION D16207.1 GI:468490
VERSION D16207.1
KEYWORDS amidase.
SOURCE Rhodococcus rhodochrous
ORGANISM Rhodococcus rhodochrous
REFERENCE 1 (bases 1 to 1958)
AUTHORS Kobayashi, M., Kameda, H., Nagasawa, T., Nishiyama, M., Horinouchi, S., Beppu, T., Yamada, H. and Shimizu, S.
TITLE Amidase coupled with low-molecular-mass nitrile hydratase from Rhodococcus rhodochrous J1. Sequencing and expression of the gene and purification and characterization of the gene product
JOURNAL Euk. J. Biochem. 217 (1), 327-336 (1993)
MEDLINE 94039054
PUBMED 7916690
REFERENCE 2 (bases 1 to 1958)
AUTHORS Kobayashi, M.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-1993) Michihiko Kobayashi, Kyoto University, Department of Agricultural Chemistry, Faculty of Agriculture, Ujiwake-cho, Kitashirakawa, Sakyo-ku, Kyoto, Kyoto 606, Japan (Tel: 075-753-6114, Fax: 075-753-6128)
COMMENT Submitted (06-May-1993) to DDBJ by: Michihiko Kobayashi
Department of Agricultural Chemistry
Faculty of Agriculture, Kyoto University
Kitashirakawa, Ujiwakecho
Sakyo-ku, Kyoto 606
Japan
Phone: 075-753-6114
Fax: 075-753-6128
FEATURES Location/Qualifiers


```
1. 1958
/organism="Rhodococcus rhodochrous"
/mol_type="genomic DNA"
/strain="J1"
/db_xref="taxon:1829"
217. 1764
/EC_number="3.5.1.4"
/codon_start=1
/transl_table=11
/product="amidase"
/protein_id="BAA03744.1"
/db_xref="GI:563884"
/translation="MSSLTPNSNQMSALNNHFRFLGTLTPELBEFAPALBATLASSET
VERLYERTAPPEQSRWTSADENPLSAMVTYTSISDEDEGLAGRTAVKDNVA
GPMWNSRTTGYTPRYDADVTRELLDAGATITKRAVCEDLFCFSGASFTSHQPVNR
PWSRSRTTGGSSGALVAGSDVMAVGDDGQIRIPAAFCIVGHKTHGLVPYT
GAFPIERTIDHLGPMWTRVSDAAMLTVLACTDGLDPROTHRIEVPVLYALAEPAAG
LRGVVTPGPTPVSDDAADVNAVTAIDVLRASGLTVEBVISIPLHLDAMAVMVIATE
GAQVMDLGNAYGNTDGFDPPELLAHFSRQLBHGHSKTVKLMGSGRYTSVTEGG
KQYAMARQLVPEVRAAYDALARYDLVWFTLYPTATKIPPTDIPDLADYLDALSMI
INTAFDVTGHPACVSPADLVHGLPTGMIIIGKHFDATVLRVAQLYHAHVNYVPVP
AAAGTLT"
ORIGIN
Query Match      12.7%; Score 200.6; DB 1; Length 1958;
Best Local Similarity 49.8%; Pred No. 5.9e-17;
Matches 649; Conservative 0; Mismatches 639; Indels 15; Gaps 5;
QY 216 ACCGACCGCGGACGAGAACCCGTTCAACGCAATTCATCCGGTCTTCGCCGGGTGGAGGCGC 275
DB 405 ACCACGCGGACGAGAACCCGCTGAGCGCCCTGGTACGTCAACCACCTCGATCAGCGAAAC 464
QY 276 CACGAGAGGGCCACTGACGACCTGACCGCGCGATCAAGCATCAAGCATCGCATCGCCATCGCGG 335
DB 465 GACGAAAGCCCCCTCGCGGGCGAACCGTCCCGTGAAAGAACACACGTCGCAGTCGCGCG 524
QY 336 TATGCCACACGAAACGGGTCCCGATGCTCCCGATGTGATCGCCACCGAGGATGCCGT 395
DB 525 CGTCCGATGATGAACCGCTCCGAAACCGTCCGAGGGCTTACCCCGCGCTACGACGCCAC 584
QY 396 GGTGTGAGGGCTGCTCGGGGAGCGCCACCATCTGTCGGCAGAGCAACCTCGAGGA 455
DB 585 CGCTGACCGGACTGCTCGAGCGCGCGCAACCATCAACCGGCAAGCGGTGTGCGAAGA 644
QY 456 CATGCCGAT--GGGTATCGGTGAAGGACGGTCTATCGGTCTCGGTGAAACCGGACAA 512
DB 645 TCTCTGCTTCTCGGCGCCAGTTTCACTTCCACCCCGCGGTTCGCAACCCCTGGGA 704
QY 513 CCCCAGCCAGCGACGGGTGATTTCCAGGGCTTCGGCGCTGCGGTGCGTCCGGCAT 572
DB 705 CGAAAGCCGCAATCACCGCGGCTCGTCCAGCGGACGCGGCGCCCTGTGTCGCGACGCGCA 764
QY 573 GGTGACTTCGCCCTGGGCGTGCATGAGCAGGACGATCCGGATCCCGCGCGCATGTG 632
DB 765 GTTGATATGAGTCCGCGGACGAGCGGCTTCGATCGCATCCCGCGCGGTCTG 824
QY 633 CGGATGTGTCGATGAAGGAGCCACCGGCTGTGCGGTCTTACGGCTTGACATCAT 692
DB 825 CGGCATCGTCGACACAAACCCACCGACGGAATGTGTCCTATACGGGAGCATTTCCCAT 884
QY 693 GGACACACCTTGGACCATCGGCGCATCACAGGGGGTTCGAGCTCAAGCCCGGGT 752
DB 885 CGAACGAACCATGACCATCTCGTCCGATGACGCGACGCTCAGCGACGCGCGCGCAT 944
QY 753 CTTGAGGTGTGTGGCGGGCGGACTGGCGGACCCCTCAGTGGGTGCGTAACCTTCGGA 812
DB 945 GCTACCGTCTCGCGCGGACCGAGCGGCTCGATCCCC--GACAGACCCACCGGATCGA 1001
QY 813 GCGGGAAGTACGGCTCGCGCTCGGAGGAGGATACCGGTCTGAGATCGCGGTG 872
DB 1002 ACCGGTGGACTACCTCGCGCGCTGGCCGAAACCCCGATCGGTCTGCGGTGGGTGGT 1061
QY 873 CGAGGAGTCACTGGAGCCGAAACGGTCCGACGCGGACGTGATCGCGCGGTTCACACAGG 932
```

```
1062 CACCGAAGGCTTCGACACCCCTGTCTCCGACGCTGCCGTCCGAAATGCCGTCCGACCGC 1121
933 ACTGGCGCGCTCGAGAGCGCGGTGCGACCATCAGCGGGTCTCGGTGCGCTTGTGGAC 992
1122 CATCGACGACTGCGCTGCGCGGACTTACCGTCCGAAAGTCTGATCCCCCTGGCACT 1181
993 GCGGCTTGGCTTATCCAGAGCGGCTGATGGCTTTTCAACGCGCGCTGATGGCGACTC 1052
1182 CGATGCAATGGCGCTTGAACGTGATCGCCACCGAGGAGCGGCTACCAAGATGCTCGA 1241
1053 CCGCGGTGGGTACTTCCAAAGGGGGGTGAGCTCAGCAGCGCGGTCCAGACGGC 1112
1242 CGCAATGCTTACGGATGAACACTGATGGCTTTACGATCCCGAATCATCGCCACTT 1301
1113 GSCCAGAGTCCGACCCACCAC--AAGGATCTGGCGATCTCTGTCCGGCTGATCTGGT 1169
1302 CTCCGCTCAACGACTCGAGCAGCGTCAACCACTGTGGAAGACAGTCAAACTCGTCGGG 1361
1170 GATCGGAGGACCTGCGGCGAGATACCTCGGCATCCTACGCGAGGAGCGGACGACT 1229
1362 GTCCGGCGCTACATCGGAGGTAGCGGCGGCAAGTACTACGCCATGGCGCGCAACT 1421
1230 GCGGCTGGAGCTCGGCAAGCAGATCGACGCGCTCTCCAGGACCGGGTGCATCTGTGAC 1289
1422 CGTCCCGAGTCCGCGCGCTACGACGCGCTTGGCTCGGTACGAGCTGCTGTGAT 1481
1290 CCGGACCAAGCTACCGTTGCCAAGAGCTGTTGAGCGGTGCGGCAAGACCATGTCCAT 1349
1482 GCCACCTCTCCCTACACGCGCACCAAGATCCCGAC--ACGGACATTCCTCGGTGCGCA 1538
1350 GATCCACCGATGACGGGCAATGCGATCTCAACAGTGCCTGCGGCTGGACCTCACCGGTCA 1409
1539 CTATCTGGACCGGACTGTGATGATCATCAACCGCACCATTCGAGGTACCGGTCA 1598
1410 CCGGCGCTGACGGGTGCCACCGGTCGCGGAGAGGGCTGCGCGTGGCTGCTCAAGT 1469
1599 CCGCGCTGAGTGTCCCGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1655
1470 GATAGCGCGCTTCCGAGGAGTTCGACGCTTACCGCACCGGCTGCTGCTGCTGCTGCT 1512
1656 CATCGCAAGCATTCGACGACGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1698
```

AP315580 9330 bp DNA linear BCT 27-JUL-2001
Agrobacterium tumefaciens hypothermal protein, Lrp-regulator-like protein, 1-aminocyclopropane carboxylate deaminase-like protein, Bkdr transcription regulator-like protein, enantioselective amidase (amdA), amide/urea-binding FmdB-like protein, amino acid transport LivH-like protein, and amino acid transport BraB-like protein genes, complete cds.
AF315580
AF315580.1 GI:13507180
Agrobacterium tumefaciens (Rhizobium radiobacter)
Agrobacterium tumefaciens
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
1 (bases 1 to 9330)
Trott,S., Bauer,R., Knackmuss,H.J. and Stolz,A.
Genetic and biochemical characterization of an enantioselective amidase from Agrobacterium tumefaciens strain d3
Microbiology 147 (Pt 7), 1815-1824 (2001)
21322708
PUBMED 11429459
2 (bases 1 to 9330)
Trott,S.
Direct Submission
Submitted (23-OCT-2000) Institut fuer Mikrobiologie, Universitaet Stuttgart, Allmandring 31, Stuttgart 70569, Germany

RESULT 12
AF315580
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

/mol_type="genomic DNA"
/db_xref="taxon:1829"

ORIGIN

Query Match 11.8%; Score 186.4; DB 6; Length 4775;
Best Local Similarity 47.8%; Pred. No. 3.6e-15;
Matches 627; Conservative 0; Mismatches 671; Indels 15; Gaps 5;
QY 217 CCGACCGGCGAGAGACCGGTTCAACGATTTCATCCGGTTCTGCGGGTGGAGGGCC 276
DB 534 CCAACAGCGAGGAAATCTTTGAGCGTTGGTATGTGACCAAGCATCCGCGGACG 593
QY 277 ACGAGGGCCACTGAGGCACTGACCGCGCGATCAAGGACTGATGCCATCCCGGT 336
DB 594 TCGGACGGCGTCTGACCGCGCGACCGCGTGGCGATCAAGGCAACGTCGCGTGGCCGGA 653
QY 337 ATGCCACCAACAAACGGGTCGGGATGCTCCCGACTGTGTGATGCCACCGAGGATGCCGTG 396
DB 654 GTTCCGATGATGAAACGGGTCTCGGACAGTAGAGGGTTCACTCCGTCTCGGACGGACT 713
QY 397 GTGGTGGAGCGGTCTGCGGACGCGGACCAATCGTCGCGCAAGAACCTCGAGGAC 456
DB 714 GTGATCACTCGACTACTGCGGCGCGGTGCAACCGTCGCGGGCAAGACTGTGTGTGAGGAC 773
QY 457 ATGGCGAT---GGTATCGTGAAGCAGCGCTACGGTCTGCGCTGCAACCGGAACAC 513
DB 774 CTGTGTTCTCGGTTCGAGCTTCAACCGGCAAGGACCGGTCCGCAATCCATGGGAC 833
QY 514 CCGGCCACCGGCAACGGGTGATCTTCCAGCGCTCCGCGCTCGCGTCCGTCGCGCATG 573
DB 834 CCACAGCGTGAAGCAGGTGATCATCCGGTGGCAGTCCGGCTCTCGTCCGAACCGTGAC 893
QY 574 GTCCGACTTCGCTCGGCTCGATGAGGACGACGATCCGATCCGATCCGCGCATGTGTC 633
DB 894 GTGATTTGGCATCGCGGGGATAGGGTGTGATCGATCCGATCCCGCGGCAATTCGTC 953
QY 634 GGACTGTGCGCATGAAGCGACCCACGCGCTGGTCCGTCTTACGGCTGATGATACATG 693
DB 954 GCGTGTGTCGACACAGCGAGTTCGCGGTCTGTCGGTATACCGGTGCAATTCCTCCATC 1013
QY 694 GACCACACTTGGACACATCGGCGCCATCACCAGGGGGTGGATCCGATCCGCGGCTGTC 753
DB 1014 GAGCGAACCAATCGACCATCTCGGCGCGATCAACGACCGGTCCAGATGCCCACTGATG 1073
QY 754 CTCGAGTGTGCGGGGCGGACTGGCGGACCTTCAGTGGGTGGTAACTTCGCGAG 813
DB 1074 CTCCTGCTATCGCGGTGCGGACGTAGACCCACGCGCAAGCGGACAGGT---CGAA 1130
QY 814 CCGGAGAACTACGGTCCGCGTTCGCGGAGGAGTATCCGGTCTGAGATCCGCTGCTGTC 873
DB 1131 GCAGGTGACTATCTGTCCACCTCGACTCCGATGTGGATGGTCTGCGAATCGGGATCGTT 1190
QY 874 GAGGAGTCACTGAGCGGAACGTGCGCGGACGTGATCCGCGTTCAACGAGGA 933
DB 1191 CGAAGAGTTTCGGGACCGGTCTCACAGCCGAGGTGACGACGATCCGCGGAGG 1250
QY 934 CTGCGCGGCTCGAGACGCGCGTTCGACCATACGCGGCTTCGCGTCCGTTGGAGG 993
DB 1251 GCACACAGTCTGTCCGAAATCGGTTCGACCGGTAGAGGAAATAAACATCCGTTGGCACTTA 1310
QY 994 GCGGCTGGCTATCCAGAGCG---GGTGTAGGCTTTCAACGCGCGCTATGCGGAC 1050
DB 1311 CATGCTTTCACATCTGGAACGTGATCGCCACGAGCGGTGTGCTTACAGATGTTGGAC 1370
QY 1051 TCGCGCGGTGGGCTACTTCCAAAGGGGCGGTGGAGCTCAGCACCCCGCTCAGCAG 1110
DB 1371 GGCAACGATACGCGATGAGCGCGGAGGTTTGTACGATCCGNNNNNNNNNNNNNNNN 1430
QY 1111 CGGCGCCAGAGTCGACCCACCAAGGATCTGGCATCTGTCCCGGTGATGCTGTG 1170
DB 1431 NNN 1490
QY 1171 ATCGCGGAGCACCTGCGCGGACGAATACCTCGGCATCCATCGGAGGCGCAGACCTG 1230

DB 1491 ACCGGCACACCGGCATCACCACTTCGCGCGGAGGTACGGCAAGCCCGAACCTC 1550
QY 1231 CGGCTGAGCTCGGCAAGCAGATCGACGCGTCTCCAGACCGGCTGCACATGCTGACC 1290
DB 1551 GTACCGTCCCGCGCGGCTAGCACTGCTTGAGACAATTCAGCTCTCTGGTGTG 1610
QY 1291 CCGACACGCGCTACCTGTCCTCAACGAGCTGTGTAGCGGTGGCAAGACACCATGTCATG 1350
DB 1611 CCCACATGCGCTACCTGCGATCCGAATACCGCGCAAGGACGCTGATCGTCAACCTTC 1670
QY 1351 ATCCCGGATGAGGGCATGCGATCCTCAACACGTCGCGCTGACCTCACCGGTGAC 1410
DB 1671 ATCAC---GAAGGCTCTCGGATGATCGCAACACAGACCGGTTCAGCTGACCGGACAT 1727
QY 1411 CCGCGCTGACGTCGCCACGCGTGGCGGAGAGGCGCTGCCCTGTGCGCTCCCAAGTG 1470
DB 1728 CCGTCCGTGCTGCTCCGCGGCTG---GTGAACGGGCTTCGGTTCGGAATGATGATC 1784
QY 1471 ATAGGCGCCACTTCGAGGATGCGCTCTACCGCACCGGCGCGCTGATCGA 1523
DB 1785 ACCGGCAAGACCTTCGACCATGCGAGCTCTCCGGTTCGGGCGCGCATTCGA 1837

RESULT 14

AE016925/c

LOCUS

DEFINITION

Chromobacterium violaceum ATCC 12472 section 16 of 16 of the

complete genome.

ACCESSION

AE016925

VERSION

AE016925.1

KEYWORDS

GI:34105514

SOURCE

Chromobacterium violaceum ATCC 12472

ORGANISM

Chromobacterium violaceum ATCC 12472

Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

Neisseriaceae; Chromobacterium.

1 (bases 1 to 208524)

Brazilian National Genome Project Consortium

The complete genome sequence of Chromobacterium violaceum reveals

remarkable and exploitable bacterial adaptability

Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)

22882880

PUBMED

14500782

REFERENCE

2 (bases 1 to 208524)

Vasconcelos A.T.R., de Almeida, D.F., Almeida, P.C., de

Almeida, L.G.P., de Almeida, R., Gonçalves, J.A.A., Andrade, E.M.,

Antonio, R.V., Araujo, J., de Araujo, M.F.P., Filho, S.A., Azevedo, V.,

Batista, A.J., Bataus, L.A.M., Batista, J.S., Belo, A., vander Berg, C.,

Blamey, J., Bogo, M., Bonato, S., Bordignon, J., Brito, C.A.,

Brocchi, M., Burity, H.A., Camargo, A.A., Cardoso, D.D.P.,

Carneiro, N.P., Carraro, D.M., Carvalho, C.M.B., Cascardo, J.C.M.,

Cavada, B.S., Chueire, L.M.O., Pasa, T.B.C., Duran, N., Fagundes, N.,

Falcao, C.L., Fantinatti, F., Farias, I.P., Felipe, M.S.S.,

Ferrari, L.P., Ferro, J.A., Ferro, M.I.T., Franco, G.R.,

Freitas, N.S.A., Furlan, L.R., Gazzinelli, R.T., Gomes, E.A.,

Goncalves, P.R., Grangeiro, T.B., Grattapaglia, D., Grissard, E.C.,

Guimaraes, C.T., Hanna, E.S., Hungria, M., Jardim, S.N., Laurino, J.,

Leoi, L.C.T., Fassarella, L., Lima, A., Loureiro, M.F., Lyra, M.C.P.,

Macedo, M., Madeira, H.M.F., Manfro, G.P., Maranhao, A.Q.,

Martins, W.S., de Mauro, S.M.Z., de Medeiros, S.R.B., Meissner, R.V.,

Menck, C.F.M., Moreira, M.A.M., Nascimento, F.F., Nicolas, M.F.,

Oliveira, J.G., Oliveira, S.C., Paixao, R.F.C., Parente, J.A.,

Pedrosa, F.O., Pena, S.J.D., Pereira, J.O., Potrich, D.P.,

Pinto, L.S.R.C., Pinto, L.S., Porto, J.I.R., Rondinelli, E., dos

Neto, C.E.R., Reis, A.M.M., Rigo, L.U., Rondinelli, E., dos

Santos, E.B.P., Santos, F.R., Schneider, M.P.C., Seuneh, H.N.,

Silva, A.M.R., da Silva, A.L.C., Silva, D.W., Souza, E.M.,

Souza, R.C., Steffens, M.B.R., Steindel, M., Teixeira, S.R.,

Urmenyi, T., Vettore, A., Wassem, R., Zaha, A. and Simpson, A.J.G.

Direct Submission

Submitted (22-JAN-2003) Labinfo, LNCC - Laboratorio Nacional de

```

Computacao Cientifica, Rua Getulio Vargas 333, Petropolis, RJ
25651070, Brazil
FEATURES
    source
        1..208524
            /organism="Chromobacterium violaceum ATCC 12472"
            /note_type="genomic DNA"
            /strain="ATCC 12472"
            /db_xref="taxon:243365"
            complement(96..563)
            /gene="pilB2"
            /locus_tag="CV4209"
            complement(96..563)
            /gene="pilB2"
            /locus_tag="CV4209"
            /note="identified by sequence similarity; putative; ORF
            located using GeneMark/Blasx/COG2165"
            /codon_start=1
            /transl_table=11
            /product="type IV pilin"
            /protein_id="AAQ61869.1"
            /db_xref="GI:34105515"
            /translation="MKQRKQGGFTLIELMIVAVIGILAAIAPAYQDYTKRVRSE
            GLADAAKTATYATYATNNAWAGTAPFNTYGLAAITGNSVSSINVLAVGVQIA
            YNSTVNGALVDLVPVTVSAGSVQVYCTGTGTTAIASQLSANVPVSTCRQ"
            838..2184
            /gene="bioA"
            /locus_tag="CV4210"
            838..2184
            /gene="bioA"
            /locus_tag="CV4210"
            /EC_number="2.6.1.62"
            /note="identified by sequence similarity; putative; ORF
            located using GeneMark/Blasx/COG0161"
            /codon_start=1
            /transl_table=11
            /product="adenosylmethionine-8-amino-7-oxononanoate
            transaminase"
            /protein_id="AAQ61870.1"
            /db_xref="GI:34105516"
            /translation="MSNOAWLERSFAAVWHPTOMKHERLPIVIPIASADQVMTDPD
            GKRYLDGVSSVMNLFHGHPRIKQVPRKLDLSLHVMLAGPHTRPVLSERLAALS
            GLGHAFYSGDGSATEIALKMSFHYKXNVGPQKTRFVLSLNSVHGTVGALAVTDVP
            LFSSTYABLLPGRAPSPDARQAVPGERAADVALRAARSENLLAKHGGEIAAIVE
            PLVQAAGAHDPFIYELRRLCDQYQVHLIADIEIAVFGRTGSGFPAYQQAGITPDF
            LCLSGITGTFPLPSVLTREDEIYQAFYHDDVARGFLHSGTGNALCAALAVLDI
            FDEEDVNAKAEQVTRQWRPLAERPERIFRQGMWAFVDVTARSPALAYFSA
            MLNRGLLRPTGKSVYFPPYTLKSEEMDLASAVDALDEVLAGSVGGDPTALP"
            2347..3816
            /locus_tag="CV4211"
            2347..3816
            /locus_tag="CV4211"
            /note="identified by sequence similarity; putative; ORF
            located using Glimmer/Blasx"
            /codon_start=1
            /transl_table=11
            /product="conserved hypothetical protein"
            /protein_id="AAQ61871.1"
            /db_xref="GI:34105517"
            /translation="MRQARPGCLVPAGPRTDGSRAALQAAQLSEGAPLLSAGERS
            FARRRGLAANAAGHQHQAQVQVPGQLQANPARKILPPFAAGATAQAGAAVHDD
            GGAATPEHQPAKLLISAARLLAGLPVAVAGAAARLAGAVAGRRRHAGRHLPAAASAR
            SDRHQVAPRRPAIGQATGLRPGATGDAVAGGPDGPGGLCAGPGRCGCPAALSAGOA
            IAVPARLPARTGRAGDETCPTTADQDGRSLSGABERESPAEPSPGSGMAPPT
            CACPLGDRGGDHRRRRHLPQRQLAARRRRPRRQRPAPAPARLLPADCRQPE
            RNLPPAPARALARRGAVAGPARPAGRVSVRSALGIDIALRQSGNRRRDAGP
            AAAAGEHAQHQHPQRQPTGLAARQEPAGAAVAPSPQLQPARVQASGRKRATG
            PRFQAASAIPPPFPADGIPPPQRAILLTWKRRRFAFSVTLPSL"
            3845..4963
            /gene="proB"
            /locus_tag="CV4212"
            3845..4963
            /gene="proB"
            /locus_tag="CV4212"

```

```

/EC_number="2.7.2.11"
/note="identified by sequence similarity; putative; ORF
located using Glimmer/GeneMark/Blasx/COG0263"
/codon_start=1
/transl_table=11
/product="glutamate 5-kinase"
/protein_id="AAQ61872.1"
/db_xref="GI:34105518"
/translation="MHSVIHASDRIVVVKVGSLSVTDGRGLDLNALARWAERVAELAKR
RGKQVLYSSGAIAEGCORLWTPKGVHQLQAAAAGQMGCLCOAYESAPRSFGLRT
AOILLTHEDLADTRYLNARSLTLLNLNVVPIINENDTVATSIIRFGDNDTLGALV
TNLIADALVILTDQGLYSADPRKHPDAEPIHQAEAGDELEDNAGAGSGSVGTGGM
ITYKILAAKRAARSGAATVASGRPEHLSRLDGEAIGTQLVAVACVDAARQKWLADH
LKLARGLLDDGAALAIERGTSLLPVGSVADGDFLRGEAVACVDAARQKWLADH
YSSDEARQIMRKSTREIEAALGYIVEPELIHRDNNVAL"
5057..6130
/gene="potF1"
/locus_tag="CV4213"
5057..6130
/gene="potF1"
/locus_tag="CV4213"
/note="identified by sequence similarity; putative; ORF
located using
Glimmer/GeneMark/Blasx/COG0687/TC:3.A.1.11.2"
/codon_start=1
/transl_table=11
/product="polyamine ABC transporter system,
substrate-binding protein"
/protein_id="AAQ61873.1"
/db_xref="GI:34105519"
/translation="NNKIATAALAAVACGSAAEVDLHIYNNNALSPDTAKRFQSC
KCKLVQDYVDNEEMAKIAGAGYDWFPTAFVATVLLKQGLQPLDKSLPLNWK
LNSGVALNQDPBPGNRVAAPTVVSLLTIGYNATOLKAGVEKGSNALVDFDPAALLA
KTKGVVTLDSORELMAALMYLGHKDNSTPADMKAAETIRKAKPYAAFNQSYI
KELTVNHWALGYSNDLFOAQODAKNAKRFPELAYRQOEGNVLAIDNMILKDAPR
POLAHKFINFMDLGRNASEISIGATNPVKAEEAFFRQIKANPVIMLDPSKGIYA
LKLDVYKRRRLNLSQVKIGR"
complement(6173..6757)
/locus_tag="CV4214"
6173..6757
/locus_tag="CV4214"
/note="identified by sequence similarity; putative; ORF
located using Glimmer/GeneMark/Blasx/COG3575"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAQ61874.1"
/db_xref="GI:34105520"
/translation="WPSGRLEQQLSGLVSLWALRAARALGLSDWCIGAGVR
GLVWDALHGFASEPSALSDDLVIYDAENLSPDREALQORLASVHPDLFWHTVQAA
VHLWFADRFGYVPELPSLEAALATWPEYATAVGVNALNADDSLRITIAFPGLDILFAMT
VRRNPARASLEIYRQVAKRYQARMPRTVVLE"
complement(7321..8358)
/gene="ruvB"
/locus_tag="CV4215"
complement(7321..8358)
/gene="ruvB"
/locus_tag="CV4215"
/note="identified by sequence similarity; putative; ORF
located using Blasx"
/codon_start=1
/transl_table=11
/product="holliday junction DNA helicase ruvB"
/protein_id="AAQ61875.1"
/db_xref="GI:34105521"
/translation="MIETDKLFGAAPERRIVTPQASOEERALERALPKLIDEVYVGQ
KXARQELEIFEAKKRGALDVLFGPGKTKTTLAHIVAREGVNLRQTSFVLE
RAGDAAALTNLEPHDVLFIDEIHRILSPVERILYPALEDYQIDIMIGESPAARVYKI
DLPPTLVGATAGTWTNPRFGIVRFBYNAEELTRIVSAGLNVQLSDDG
AREVAKRSRGTBRIANRLRRVRYDAEVKSDGVTWAVADALAMLDVDPAGLDVMDR
KLQAILEKPSGPGVLDNVAAGISTDTIEDVIEPFLIQOGLQRTFRGRMATAQA
YLHFLGUPVKDA"
8726..9358

```

```
/gene="ribB"
/locus_tag="CV4216"
8726..9358
/gene="ribB"
/locus_tag="CV4216"
/notes="identified by sequence similarity; putative; ORF
located using GeneMark/BLASTx/COG0108"

Query Match      11.8%; Score 186; DB 1; Length 208524;
Best Local Similarity 49.2%; Pred. No. 1.4e-15; Indels 30; Gaps 5;
Matches 650; Conservative 0; Mismatches 640;

QY 183 GAGCGCGTACAGGACCGGACATCGCGCGGAAACCGACCGGCGAGAGACCGGTTCAA 242
DB 154426 GCGCTGAACCGCAGCTCAACGCCATCGTACCGTAGACCGGAAAGACGCTGCGCA 154367
QY 243 CGCATTTATCCGGTTCTGCGGGTGAAGGCGCCAGGAGGGGCACTAGAGCACTGAC 302
DB 154366 GCGCGCGGCGCGACCGCGCTCGCGCGGCGATGCGCGCTGACCGGCGTGC 154307
QY 303 CGCGCGGATCAAGGATCGATCGCCATCGCGGTATGCCACACGAAAGGGTCCCGGAT 362
DB 154306 GCTGGTGCAAGGACCTGTTCTGCCAGCAAGTTGGAACAACAGCTGCGCTCCAGAT 154247
QY 363 GCTCCCGACTGTATCGCCACCGAGATGCGGTGGTGGAGCGGCTGCTCGGCGCAGG 422
DB 154246 GCTGGAACAATTCGTTTGGCCGTACAGCGCCACGTTGGTAGACATGCGCGCGCGG 154187
QY 423 CGCCACCATCGTCGGAAGACGAACTCGAGGACATGGGTATCGG--TGAAGG 479
DB 154186 CATGTTGACGCTGGCGCGCCAAATATGACGAGTTTCGGATGGGCTCGTGAATGAAA 154127
QY 480 CAGCGTTACGGTCTGCGCTGAACCGCAACACCGCGCCAGCGGCAACGGGTGGATCTTC 539
DB 154126 CTGCTTTACGGCGCGGTGAAGAACCGGTGGGACCTGAACCGGATTCGGGCGGTTCTGTC 154067
QY 540 CAGCGGCTCGCGGCTCGCGTCTGCTCGCGCATGCTCGCTCGCGCTCGCGGTGATCA 599
DB 154066 CGCGGCTCGCGGCTCGCGTCTGCGCGCATGCTCGCGCTCGCGGTGATCA 599
QY 600 GGAGGCGAGATCGGATCGCGATCGCGCGGATCGGTGCGGATCGGATGAAGCGGACCA 659
DB 154006 CGCGGTTTCCATCGCGCGCGGCGGACGATTCGGGCTGACCGGATCAAGCGGACCTA 153947
QY 660 CGCGCTGCTCGCTCTACGGCTGACATACATGAGACCAACCTTGGACCAATCGGGC 719
DB 153946 CGCGGTTGTTGCGGTTACGGCATGTTGGCTTACGCTTCTCGCTGACCGGCGGCGC 153887
QY 720 CATCACAGGCGGCTGAGCTCAACCGCGGCTCTCGAGGTTGTTGGCGGCGGCGGACTG 779
DB 153886 GATCGCCACAGCGCGGAGCTGCGCGCTGATGCTGAACGTGATGCGCGGCTTCGACGC 153827
QY 780 GCGGACCTCAGTGGTGGTGAACCTTCGAGGCGGAGAACTAGCGTCCGCGCTCGG 839
DB 153826 GCGGACTCCACGAGCTG-----GAGCGCGCAAGGAGAGACTAGCGCGGACCTGAA 153773
QY 840 CGAGGAGTATCGGTTGAGATTGCGGCTGCTGAGGAGTCACTGAGCGGAAACGGTGC 899
DB 153772 CCAGTCTTGTCCGACTGAGGTTGGCTTCCCAAGGATATCTTGGCGCGGCTCGGA 153713
QY 900 GACCGCGGATCGGATCGCGGTTCAACAGGAGTCTGGCGGCGCTCGAGAGCGCGGTCG 959
DB 153712 CGCGGAGCTCGCGGCGCGCTGCACAAACGCGGCTGGCGAGCTGAAAGAGTGGCGCGCA 153653
QY 960 GACCATCGAGCGGCTCGGTGCGGTTGTGAGCGGCGGCTGCGCTATCCAGAGCGGCT 1019
DB 153652 GCGGCTGAGATAGCTGCGGCTGACAGGAGTGTGATCCCGGCTTATTCGTGATCGC 153593
QY 1020 GATGGTTTCAACCGCGCGCTATGGCGGATCGCGCGGTGTTGGGCTACTTCCACAGGG 1079
DB 153592 CCCGCGGAGGCTCGGACCAACCTGTCCGCTACGACGGGCTCGGCTACGCGCCAC--CG 153536
QY 1080 GCGGTGAGCTGACGACCGCGCTCAGACGCGGCGGCGGAGTCTGCACACCCACAGGA 1139
```

```
DB 153535 GCGCAAGGACTACAGGACTTGTTCGACATGTACGAGAAGACCGCGCCGAGGCTTCGG 153476
QY 1140 TCTGGGATCTGTCCCGGCTGATGTGGTGATCGCGGAGACACTGTCGCGAGCAATACCT 1199
DB 153475 CGACGAGGTCAAGCGCGCATCTCTGTGGGTAGTATGTG---CTGAGCCACGGTTACTA 153419
QY 1300 GGGCATCCACTACGGAAGCGGAGAACCTCGCGCTGAGCTGGCAAGCAGATCGACGC 1259
DB 153418 CGAGCCTACTACTCTGAAGGCGGAGAGATCCGCCGCTGATCGCAACGACTTCAAGGC 153359
QY 1260 CGTCTCCAGSACCGGCTGCACTGTGACCCCGACACACCTACCTTCCCAACGAGCT 1319
DB 153358 GCGCTTGGAGCAGTGCAGCTGATCTCTGGGCGCGGTGGCGCCGACCCGCTTCAATAT 153299
QY 1320 GTTGAGCGGTGCGCAAGACACCATGTCCATGATCCACGATGACGGGCAATGCGATCCT 1379
DB 153298 C-----GGCAGAAAGTCCGGGATCCGGTGAGATGATCTGTCCGACAT 153254
QY 1380 CAACACGTGCGCGCTGACCTCACCGTCAACCGCGCTGACCGTCCCAACGGGTGCGGG 1439
DB 153253 CTACACGCTGTGCGTGAACCTTGGCGGCTTGGCGGATGAGCGTGGCGCGGCTTCGC 153194
QY 1440 CGAGAAGGCTGCGCGCTTGGCTTCCAGTATAGGCGGCGCACTTCGAGGAGTCAAGCT 1499
DB 153193 GCGCAAGCGCGCGCTGATCGGCTTGCAGATCATCGCAACTACTTTCGCGAAGCCAGAT 153134

RESULT 15
BD061400 2822 bp DNA linear PAT 27-AUG-2002
LOCUS Rhodococcus bacterium-derived nitrile hydratase gene and amittase
DEFINITION gene.
ACCESSION BD061400
VERSION BD061400.1 GI:22607005
KEYWORDS JP 2001292772-A/2.
SOURCE Rhodococcus sp.
ORGANISM Rhodococcus sp.
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Nocardiaceae; Rhodococcus.
REFERENCE 1 (bases 1 to 2822)
AUTHORS Kamaike, H. and Aoki, Y.
TITLE Rhodococcus bacterium-derived nitrile hydratase gene and amittase
JOURNAL Patent: JP 2001292772-A 2 23-OCT-2001;
SHOWA DENKO KK
OS Rhodococcus sp.
FN JP 2001292772-A/2
PD 23-OCT-2001
PF 10-APR-2000 JP 2000107855
PI HARUMI KAMAIKE, YUJI AOKI
PC C12N15/09, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/80, C12N9/
PC 88, C12P13/00,
PC C12P13/02// (C12N15/09, C12R1.01), (C12N9/80, C12R1.01), (C12N9/88,
PC C12R1.01),
PC C12N15/00, C12N5/00, (C12N15/00, C12R1.01)
CC amittase
CC amittase
PH Key Location/Qualifiers
FT CDS Location/Qualifiers
1..2822 (1094)..(2491).
/organism="Rhodococcus sp."
/mol_type="genomic DNA"
/db_xref="taxon:1831"

FEATURES
source
1..2822
Query Match 11.5%; Score 182.2; DB 6; Length 2822;
Best Local Similarity 53.4%; Pred. No. 1.5e-14; Indels 6; Gaps 2;
Matches 429; Conservative 0; Mismatches 369;
QY 216 ACCGACCGGCGAGAGACCGGTTCAACGCACTTCACTCGGTTCTCGCGGTGGAGGCGC 275
DB 1282 ACCGACCGGCGAGAGACCGGTTGAGCGCTGCTGACCTGACCTGCTGACGAGAAC 1341
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 03:11:33 ; Search time 680.85 Seconds
(without alignments)
9846.004 Million cell updates/sec

Title: US-10-658-691-4

Perfect score: 1578

Sequence: 1 gtgcaccacatcgccatt.....ggcaccgtgtggtggtag 1578

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_29Jan04.*

1: Geneseqn1980s.*

2: Geneseqn1990s.*

3: Geneseqn2000s.*

4: Geneseqn2001as.*

5: Geneseqn2001bs.*

6: Geneseqn2002as.*

7: Geneseqn2003as.*

8: Geneseqn2003bs.*

9: Geneseqn2003cs.*

10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	182.2	11.5	2822	4 AAH21845	Aah21845 Rhodococc
2	180.8	11.5	1879	2 AAQ11991	Aaq11991 Enantiose
3	180.8	11.5	1879	2 AAQ25616	Aaq25616 Enantiose
4	175.4	11.1	1566	2 AAT82385	Aat82385 Rhodococc
5	175.2	11.1	1479	7 ACA37683	Ac37683 Prokaryot
6	172.8	11.0	110000	4 AAI99683_33	Continuation (34 o
7	172	10.9	1482	7 ACA38250	Ac38250 Prokaryot
8	170.4	10.8	1482	7 ABZ71125	Abz71125 Mycobacte
9	170.4	10.8	1485	7 ACA40772	Ac40772 Prokaryot
10	170.4	10.8	110000	4 AAI99682_33	Continuation (34 o
11	157.4	10.0	4730	2 AAT82387	Aat82387 R. rhodoc
12	152.2	9.6	1488	7 ACA26208	Ac26208 Prokaryot
13	151	9.6	1455	4 AAS4287	Aas4287 Pseudomon
14	151	9.6	1455	7 ACA42618	Ac42618 Prokaryot
15	146.4	9.3	349980	6 ABO81844	Abq81844 Bifidobac
16	141.4	9.0	1488	7 ACA23486	Ac23486 Prokaryot
17	141.4	9.0	1536	7 ACA27226	Ac27226 Prokaryot
18	140.6	8.9	1449	7 ACN43522	Ac43522 Prokaryot
19	133.2	8.4	2438	2 AAT47765	Aat47765 Comamonas
20	130.6	8.3	1038	7 ADA71204	Ada71204 Rice gene
21	125.4	7.9	1398	9 ADE50885	Ade50885 V. parado
22	124.2	7.9	1485	7 ACA25570	Ac25570 Prokaryot
23	123.8	7.8	1449	7 ACA45943	Ac45943 Prokaryot

24	123.4	7.8	110000	4 AAI99682_26	Continuation (27 o
25	123.4	7.8	110000	4 AAI99683_26	Continuation (27 o
26	121	7.7	1494	7 ACA39866	Ac39866 Prokaryot
27	120	7.6	2456	2 AAQ13362	Aaq13362 Gene enco
28	119.4	7.6	1521	7 ACA52610	Ac52610 Prokaryot
29	114.2	7.2	1443	7 ABZ39903	Abz39903 N. gonorr
30	114.2	7.2	1443	7 ACA41226	Ac41226 Prokaryot
31	114.2	7.2	1458	6 ABK73873	Abk73873 Bacillus
32	113.4	7.2	1386	7 ABZ71160	Abz71160 S. muraya
33	113.4	7.2	1446	7 ACA41891	Ac41891 Prokaryot
34	113.4	7.2	36321	7 ABZ71131	Abz71131 Streptomy
35	111.6	7.1	50925	3 AA81487	Aa81487 N. mening
36	111.6	7.1	110000	3 AA81490_13	Continuation (14 o
37	111.6	7.1	349980	3 AAF21610	Aaf21610 Neisseria
38	111	7.0	861	4 AAH27159	Aah27159 Amidase D
39	110.4	7.0	23673	6 ABZ75344	Abz75344 Human R11
40	108.8	6.9	1844	2 AAO38949	Aao38949 C. acidov
41	107.8	6.8	22934	4 AAS59613	Aas59613 Propionib
42	107.8	6.8	22934	7 ACF64542	Acf64542 Propionib
43	106.8	6.8	1395	7 ACA31528	Ac31528 Prokaryot
44	106.6	6.8	7127	6 ABL58898	Ab158898 Ralstonia
45	102	6.5	1621	9 ADD13197	Add13197 C. glutam

ALIGNMENTS

RESULT 1
AAH21845
ID AAH21845 standard; DNA; 2822 BP.
XX
AC AAH21845;
XX
DT 15-AUG-2001 (first entry)
XX
DE Rhodococcus sp. amidase encoding DNA SEQ ID NO:6.
XX
KW Rhodococcus; nitrilase; nitrile hydratase beta subunit; amidase;
KW nitrile hydratase alpha subunit; carboxylic acid; cyanocarboxylic acid;
KW Pharmaceutical; agricultural drug; dye; ds.
XX
OS Rhodococcus sp.
XX
FH Key
CDS Location/Qualifiers
FT 1094..2491
FT /*tag= a
FT /product= "amidase"

WO200130994-A1.
XX
PD 03-MAY-2001.
XX
PF 25-OCT-2000; 2000WO-JP007464.
XX
PR 26-OCT-1999; 99JP-00303212.
PR 26-JAN-2000; 2000JP-00021797.
PR 10-APR-2000; 2000JP-00107855.
(SHOW) SHOWA DENKO KK.
XX
PI Aoki H, Kamachi H;
XX
DR WPI; 2001-308634/32.
DR P-PSDB; AAB98070.
XX
PT Producing carboxylic acids using Rhodococcus to convert cyano groups of
nitrite compounds to carboxyl groups.
XX
PS Claim 29; Page 75-80; 89pp; Japanese.
XX
CC The present invention describes a method for producing carboxylic acids,
comprising using a microorganism, whose ability to convert cyano groups
to amide groups has been removed or reduced, to convert at least one

cyano group of a nitrile compound to a carboxyl group. Also described are: (1) a microorganism for use in the novel method; (2) a plasmid containing DNA for Rhodococcus nitrilase gene; (3) a host transformed by the plasmid of (2); (4) a nitrilase gene, comprising a 1531 base pair sequence (AAH21843), and encoding a 366 residue amino acid sequence (AAB98067); (5) producing nitrilase, comprising culturing the host of (3) under expression conditions, and recovering the polypeptide; (6) a nitrile hydratase gene, comprising a 2822 base pair sequence (AAH21844), encoding a 229 nitrile hydratase beta subunit (AAB98068) and a nitrile hydratase alpha subunit (AAB98069); (7) producing amides using hosts transformed by a 2822 base pair sequence (AAH21845), encoding for Rhodococcus amidase (AAB98070). The method can be used for producing carboxylic acids and cyanocarboxylic acids which are useful in pharmaceuticals, agricultural drugs, dyes and as starting materials for other chemicals

XX
SQ Sequence 2822 BP; 478 A; 1034 C; 869 G; 438 T; 0 U; 3 Other;

Query Match 11.5%; Score 182.2; DB 4; Length 2822;
Best Local Similarity 53.4%; Pred. No. 1.6e-22;
Matches 429; Conservative 0; Mismatches 368; Indels 6; Gaps 2;

QY 216 ACCGACCGCGCGAAGACCGGTTCAACGATTATCCCGTTCTGCGGGTGGAGGGCG 275
DB 1282 ACCACCGCGGACGAGAACCGGTGAGCGCTGGTAGCTCACCACCTCGATCAGCGAAC 1341

QY 276 CACGGAGGGCCACTAGCGACTGACCGCGGATCAAGAGTCAAGTCAATGCCATCGCCGG 335
DB 1342 CGAAGAGGGCCCTCGCGGGGGAACGCTGCGCGTGAAGACAACTGCGAGTGCCTGG 1401

QY 336 TATGCCACACCAAGCGGTCCTCGGATGCTCCGACTGTCGATCGCCACCGAGGATCGGT 395
DB 1402 CGTGCCGATGATGAACGGCTCCGACCTCGAGGCTTCACCCCGCTACGAGCCAC 1461

QY 396 GTGTGTGAGCGCTCTCGCGCAGCGCGCCACCATGCTGCGCAAGAACCTCGAGGA 455
DB 1462 CGTCGTACCGCGACTCTCGACCGCGCGCAACCATCACCAGCAAGCGCTGTGCGAAG 1521

QY 456 CATGGCGAT--GGGTATCGGTGAAGCAGCGCTACGGTCTCGCTGAGAACCCGACAA 512
DB 1522 TCTCTGCTCTCGCGCGCAGCTTCACTTCCACCCCGAGCGCTCCGACCCCTGGGA 1581

QY 513 CCCCAGCGCGCAGCGCGTGTGATCTTCCAGCGGCTCCGCGTGCCTGCGTGCCTGGCAT 572
DB 1582 CGAAGCGCGCATCACCGCGGCTGCTCCAGCGCAGCGCGCTGTCGCCAGCGGCA 1641

QY 573 GFTCGACTTCGCTCGCGGTCGATGAGCAGCAGCATCGGATCCGCGCGCATGTG 632
DB 1642 GTGTGATATGGAGTCCGCGCGCACCGAGCGGCTTCGATCCGATCCCGCGCGTTCG 1701

QY 633 CGGACTGGTGGCATGAAGCGACCCACCGCGCTGTGCTGTCTTACGGCTGACATACAT 692
DB 1702 CGGCATCGTCGGACACAAACCCACCGACTGTGTCCTTATACGGAGCATTTCCCAT 1761

QY 693 GGACACACTTGGACACATCGGCGCCATCACAGGGGGTTCAGTCAAGCGCGGT 752
DB 1762 CGAACGAACATCGACCACTCGGTCCGATGACGCGACGTCAGCAGCGCGCGCAT 1821

QY 753 CTTCGAGGTGTTGGCGGGCGCGACTGGCGCGACCTCAGTGGTGGCTAACTTCCGGA 812
DB 1822 GCTCACCGTCTCGCGCGCACCGAGCGCTCGATCCCC--GACAGACCCACCGATCGA 1878

QY 813 CGCGGAGACTACGCTCGCGCTCGCGAGGAGATTCGGTCTGAGATTCGCGGTGCT 872
DB 1879 ACCGTGGACTACCTCGCGCGCTGGCGGAACCGCATCGGTCTCGCGGTGGTGGT 1938

QY 873 CGAGGAGTCACTGGAGCCGAAACGGTTCGACCGCGGACGTGATCGCGGTTCACACGAG 932
DB 1939 CACCGAAGGTTTCGACACCCCTGTCTCCAGCGCTGCCGTCGACATGCGGTGCGACCGC 1998

QY 933 ACTGGCGGCTCGAGAGCGCGGTGCGACCATCGAGCGGTCTCGGTGCTTGTGAC 992
DB 1999 CATCGCGTACTCGGTTCGCGCGGACTTACCGTTCGAAGAGGTCTCGATCCCTCGCACCT 2058

QY 993 GCGCGCTGCGCTATCCAGAGCG 1015
DB 2059 CGATCGATGCGCTCTGGAACG 2081

RESULT 2
ID AAQ11991 standard; DNA; 1879 BP.
XX AAQ11991;
AC AAQ11991;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 30-AUG-1991 (first entry)
XX
DE Enantioselective amidase gene of Brevibacterium.
XX
XX ESA; S(+)-ketoprofen; racemase; ss.
XX
OS Brevibacterium sp.
XX
FH Key Location/Qualifiers
FT RBS 230..234
FT /*tag= b
FT /*standard_name= "Shine-Dalgarno sequence"
FT CDS 245..1810
FT /*tag= a
FT /*product= "enantioselective amidase"
XX
XX EP433117-A.
XX
XX 19-JUN-1991.
XX
XX 15-NOV-1990; 90EP-00403232.
XX
XX 11-DEC-1989; 89FR-00016332.
XX
XX (RHON) RHONP-POULENC SANTE.
XX (PETR)/ PETRE D.
XX
XX Petre D, Cerebaleud E, Mayaux JF, Yeh P;
XX WPI; 1991-179908/25.
XX
XX DNA encoding enantio-selective amidase - used for prodn. of the enzyme
XX for enantio-selective hydrolysis of amide(s) to acids.
XX
XX Claim 2; Fig 8; 36pp; English.

Probe Sq762 (AAQ11989) and probe Sq918 (AAQ11990) were used to screen genomic Brevibacterium DNA. Both probes hybridised strongly to a 5.4kb PstI fragment. Cloning and restriction analysis of this fragment showed that the gene coding for the enantioselective amidase is located on a 2.3kb BamHI-PstI fragment (oriented in the direction BamHI towards PstI). The fragment was sequenced. See also AAQ11988 and AAQ11992. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 27-AUG-2003 to correct OS field.)

SQ Sequence 1879 BP; 409 A; 612 C; 531 G; 327 T; 0 U; 0 Other;

Query Match 11.5%; Score 180.8; DB 2; Length 1879;
Best Local Similarity 48.1%; Pred. No. 2.8e-22;
Matches 704; Conservative 0; Mismatches 742; Indels 18; Gaps 6;

QY 66 CGCCCGCCACACTTCGACCTCGACGAGGAACTGGCGCCCGACCTCGTTCGGTCTGGC 125
DB 286 CGCAGGCGATTACGGCATCACTCTGCACAAACAGCCCGCTCGAGTGGCGGCACTGAT 345
QY 126 GGAGATGCTGACTGCTTTCGACTGATCGACGAACTACGCAACCCCGCGCGCGAC 185
DB 346 CGACGGAGCACTGGGTCTCTACGAGTCTGTCGACCGAGTGTGACCGGCGACCCC 405

186 GCCTACACGACCGGACATCGCGCGGAAACCGACCGGCGAGACCCGCTCAACGC 245
187 |||||
406 GCCGACCAAGTACGAGGACG---CGGTCCAGTGGAGGAAATCTTTGAGCGC 462
407 |||||
246 ATTATCCCGTTTCGCGGTGGAGGCGCCACGAGGGGCCACTGAGCGACTGACCGC 305
408 |||||
463 TTGGTATGTGACCAACCAAGATCCGCGGACGTCGAGCGCGCTCTGACGGCGGACGCT 522
409 |||||
306 CGGATACAGGACTGATCCCATCCCGGTATGCCACCAACGAGCGGTCCGGATGCT 365
410 |||||
523 GGCATCAAGGACAAGTGAACCTGGCGGAGTTCGGATGATGAACGGATCTCGACCGT 582
411 |||||
366 CCGACTGTGATCCCAACAGAGATCCGCTGTGTGTGAGCGGCTCTCGCGCGAGCGC 425
412 |||||
583 AGAGGATTTACTCCGTCACGACGCGACTGTGGTCACTCGACTACTGCGCGCGGTGC 642
413 |||||
426 CACCATCGTCGGAACAGNACCTCAGGACATGGCGAT---GGGTATCGGTGAAGGAG 482
414 |||||
643 AACCGTCGCGGGAAGAGCTGTGTGAGGACCTGTGTTCTCGGTTCGAGCTTCACAC 702
415 |||||
483 CGTCTACGCTCTCGCTGAACCCGGAACACCCCGCCGCGACCGGTGGATCTTCCAG 542
416 |||||
703 GGCAGCGGACCGTCCGCAATCCATGGNACCGGACGCGGAGCGAGTGGATCATCCG 762
417 |||||
543 CGCTCCGCGCTGCGCTGCTGCGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
418 |||||
763 CGGAGTGCAGCACTCTGTCGCAACCGTGTGACGCTGCTGCTGCTGCTGCTGCTG 822
419 |||||
603 AGGACGATCCGATCCGCGCGCATGCTGCGGACTGCTGCGGATCAAGCGGACCCAG 662
420 |||||
823 CGATCGATCCGATCCGCGCGCATGCTGCGGCTGCTGCGGATCAAGCGGACCCAG 882
421 |||||
663 CTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722
422 |||||
883 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 942
423 |||||
723 CACAGGGGCTGAGCTCAACCGCGGCTCTGAGGTGCTGCGGCGGCGGCGGCTGCG 782
424 |||||
943 CACAGCGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1002
425 |||||
783 CGACCTCTGAGTGGTGAACCTTCGAGCGGAGAACTACGGCTCGCGCTCGCGGA 842
426 |||||
1003 CGACCCAC---GCCAGCCGAGTGTGGAAGCAGTGTACTTCTCCACCTCGACTC 1059
427 |||||
843 GGAAGTATCCGCTGAGATTCGGGTGCTGAGGAGTCTACTGAGCGGAGAACGGTGG 902
428 |||||
1060 CGATGTGAGCGGCTGGAATCGAATCGAATCGAATCGAATCGAATCGAATCGA 1119
429 |||||
903 GCGGAGCTGATCGCGGCTTCAACAGGAGTGGGCGGCTCGAGAGCGCGGTGCGAC 962
430 |||||
1120 GCGGAGTGCAGACCGAGTCCGCGAGCGGACACAGTCTGACCGAATCGGTGAC 1179
431 |||||
963 CATGAGCGGCTCTCGGTGCGGTGAGCGGCGGCTGCGCTTATCCAGAGCG---CGGT 1019
432 |||||
1180 GGTAGAGGAATGAATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1239
433 |||||
1020 GATGCTTTCAACGCGGCTTGGGAGTCTCGCGGTGCTGCTGCTGCTGCTGCTGCT 1079
434 |||||
1240 CAGCGCGGTGCTTACAGATGTTGGAGCGCAACGATATCGGATATCGGATGAG 1299
435 |||||
1080 GCGGTGAGCTGACGACCGCGCTCAGACGCGGCGGCGGAGTGGCGGAGGAGGAG 1139
436 |||||
1300 TTTGATGATCCGGAATGATGGACACTTTGCTTCTGACGCTATTCAGACCGCGAC 1359
437 |||||
1140 TCTGCGATCTGCTCCGCTGATGCTGTGATGCGGAGACCTGCGGAGCGAATACCT 1199
438 |||||
1360 TCTGCGAATCCGCTCAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1419
439 |||||
1200 CGGATCACTTACCGAAGCGGAGAACCTCGGCGCTGAGCTCGGCAAGCAGATCGAC 1259
440 |||||
1420 CGGCGGAGCTACGCGAAGCGGAGAACCTCGTACCGCTGCGGCGGCGGCTTACG 1479
441 |||||
1260 CGTCTCAGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1319

1480 TGCCTGAGACATTCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1539
1481 |||||
1320 GTTGAAGGCTCGGCAAGACACCATGTCCATGATCCACCGATGACGGCAATGCGATCT 1379
1321 |||||
1540 GCCGCGAAGGACGTAGATGTGCAACCTTCATCAC---CAAGGCTCTCGGATGATCGC 1596
1541 |||||
1380 CAACAGTGGCCGCTGGACCTCACCGGTGACCGCGCTGACCGTGGCCCGGCGGCGG 1439
1381 |||||
1597 CAACAGGACCATTCAGCTGACGACGACATCCGCTGCTGCTGCTGCTGCTGCTGCT 1654
1598 |||||
1440 CGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1499
1441 |||||
1655 -GTGAACGGGCTTCGGGTGCGAATGATGATCACCGGACGACACTTCGACGATGCGACGT 1713
1656 |||||
1500 CTACGACCGGCGCGCTGATCGA 1523
1501 |||||
1714 CTTCTGCTGCGACGCGCATTCGA 1737
1715 |||||
RESULT 3
AAQ25616
ID AAQ25616 standard; DNA; 1879 BP.
XX
AC AAQ25616;
XX AC
XX 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 27-NOV-1992 (first entry)
XX
DE Enantioselective amidase gene.
XX
KW Recombinant; ammonium adipate production; es.
XX
OS Brevibacterium sp.
XX
PH Key Location/Qualifiers
FT RBS 230..234
FT /*tag= b
FT /*note= "Shine Dalgarno sequence"
FT CDS 245..1810
FT /*tag= a
FT /*note= "enantioselective amidase"
XX
PN EP488916-A1.
XX
XX 03-JUN-1992.
XX
XX 28-NOV-1991; 91EP-00420422.
XX
XX 28-NOV-1990; 90FR-00014853.
XX
XX (RHON) RHONE POULENC CHIM.
XX (RHON) RHONE POULENC FIBER & RESIN INTERMEDIATES.
XX
XX Yeh P, Mayaux JF, Cerbelaud E, Petre D;
XX
XX WPI; 1992-185781/23.
XX P-PSDB; AAR24529.
XX
XX Ammonium adipate prodn. - by enzymatic hydrolysis of adipamide or
XX ammonium adipamate using amidase from Brevibacterium R312 or its
XX analogues.
XX
XX Example; Fig 9; 39pp; French.
XX
XX The sequence is that of the enantioselective amidase gene from
XX Brevibacterium R312 which was isolated using oligonucleotide probes
XX pred. corresponding to known parts of the sequence to screen genomic
XX R312 DNA which had been cleaved with restriction enzymes. The
XX enantioselective amidase it encodes can be used in the prodn. of ammonium
XX adipate by enzymatic hydrolysis of adipamide or ammonium adipamate, its
XX use provides particularly high yields. See also AAQ25617-Q25623 (Updated

CC on 25-MAR-2003 to correct PN field.) (Updated on 25-MAR-2003 to correct
CC PA field.) (Updated on 27-AUG-2003 to correct OS field.)
XX Sequence 1879 BP; 409 A; 612 C; 531 G; 327 T; 0 U; 0 Other;
SQ

Query Match 11.5%; Score 180.8; DB 2; Length 1879;
Best Local Similarity 48.1%; Pred. No. 2.8e-22;
Matches 704; Conservative 0; Mismatches 742; Indels 18; Gaps 6;
66 GCGCGGACCACTTCGACCTCGAGGAACTGGCGCCAGCTCGTTCCGGTCGTGGC 125
DB |||||
286 GCGAAGGATACCGCATCACTTCGACAAACAGCCCGGTCGAGTGGCGCGCACTGAT 345
QY |||||
126 GAGATGAGTACTGCTTCGACCTGATCGACAACTACCGAAACCGCGCAGCCGCGAC 185
DB |||||
346 GCGCGAGGACCTGGGCTCTACAGCTGCTCGACCACTGTTGACGCGGACGAGGCGACCC 405
QY |||||
186 GCGGTACAGGACCGCGACATCGCGCGGACCGACCGGACGAGACCGGTTCAACGC 245
DB |||||
406 GCGGACCACTACCGGAGCAGC---CGGTGCAAGTCGAGCGGAAATCCTTTGAGCGC 462
QY |||||
246 ATTCATCCGGTTCTCCCGGTTGAGGCGCCACGAGGGGCACTGAGCGCACTGACCGC 305
DB |||||
463 TTGATATGTGACACAGCATCCCGCGACGTCGAGCGGCTCTGACCGCGCGACGCT 522
QY |||||
306 CGCGATCAAGGACTGATCGCATCGCGGTATGCCACCGACGACGCGTCCCGATGCT 365
DB |||||
523 GCGGATCAAGGACAACTGACCGTGGCGGAGTTCGATGATGAAACGATCTCGGACGGT 582
QY |||||
366 CGCGACTGTGATCGCACCGAGATGCGGTGTGTGAGGCGGTGCTCGCGGCGAGCGC 425
DB |||||
583 AGAGGATTTACTCCGTACGCGGACGACGCTGTGTCACTCGACTACTGGCGCGGTGC 642
QY |||||
426 CACCATCTCGCGACAGCACTCGAGGACATGCGGT---GGGTATCGGTGAAGGAG 482
DB |||||
643 AACCGTTCGCGGCAAGCTGTGTGAGGACCTGTGTTCTCCGGTTCGAGCTTCACACC 702
QY |||||
483 CGTCTACGCTCTGCGCTGAACCCGAAACCCCGCCACCGCACTGGTGTATCTTCAG 542
DB |||||
703 GCGAAGCGACCGGTCCGCAATCCATGGACCGCGGACGAGGATGATCATCGG 762
QY |||||
543 CGGCTCCGCGCTGCGGTGCGGATGTCGACTTGGCCCTGGGCGGTGATGAGGC 602
DB |||||
763 CGGCACTGACGACTCGTCGCAACCGGTGACGCTGATTTGCAATCGCGGCGGATCAAG 822
QY |||||
603 AGCGAGCATCCGATCCCGCGCATGTCGCGACTGTCGATGAGGAGGAGGACCGCG 662
DB |||||
823 CGGATCGATCCGATCCCGCGCATGTCGCGCTGTCGCGGACAAAGCGCACTTCGG 882
QY |||||
663 CTGGTGCGCTTTACGGCTGACATATGAGACCACTTGGACCACTGGGCGCCAT 722
DB |||||
883 GCTCGTCCGTATACCGGTGCAATTTCCCATCGAGCGAAACATCGACCATCTCGGCGCAT 942
QY |||||
723 CACGAGGGGTCGAGCTCAACCGCGGTCTCGAGGTGTTGGCGGCGCGGCGGCGG 782
DB |||||
943 CACAGCAGGTCAGATGACGACTGATGCTCTCGGTATCGCGCGCGGCGGAGCTGAA 1002
QY |||||
783 CGACCTCAAGTGGTGGTAACTTCCGAGCGGAGAACTACGCTCGCGCTCGGCGA 842
DB |||||
1003 CGACCCAC---GCCAAGCGGACAGTGTGCAAGCAGTGACTATCTGTCCACCTCGACTC 1059
QY |||||
843 GGGATATCGGTCTGAGATTCGGGTGCTCGAGGAGTCACTGGAGCGGACGAGTGCAC 902
DB |||||
1060 CGATGTGAGCGGCTCGGAATCGGAATCGTTTCGAGAGGATTCGGGCGCGGCTCTACA 1119
QY |||||
903 GCGGAGCGTATCGCGCGCTTCAACCGAGGACTGGCGGCTTCGAGAGCGCGGTGCGAC 962
DB |||||
1120 GCGGAGGTTCGAGCGGAGTCCGCGACGCGACACAGTCTGACCGGAAATCGGTTGCAC 1179
QY |||||
963 CATCGAGCGGCTCGGTGCGGTGAGCGCGGCTGCGCTATCCAGAGG---GCGT 1019
DB |||||
1180 GGTAGAGGAAGTAAACATCCCGTGGCATCTGATCTTCCACATCTGGAACGTGATCGC 1239

QY 1020 GATGGCTTTCAACGGCGGCTATGCGGACTCCCGGCTGTGGGTACTTCCACAAGGG 1079
DB |||||
1240 CACGAGCGGTGGTCTTACAGATGTTGACCGCAACGATACGCGATGAACGCGGAGG 1299
QY |||||
1080 GCGGTGAGCGTCAGCACCGCGCTCAGACGCGGCGCCAGAGTCGACCAACCAAGGA 1139
DB |||||
1300 TTGTACGATCCGGAACGATGATGCGACATTTGCTTCTCGACGATTCAGACGCGGACG 1359
QY |||||
1140 TCTGGGATCTGTCTCCGGCTGATGCTGGTATCGGGAGCACCTGCGGCGAGGAATACCT 1199
DB |||||
1360 TCTGTCCGAAACCGTCAAACTGGTGGCCCTGACCGGCGCACACGCGCATCACACCTCG 1419
QY |||||
1200 CGGATTCATCGGAGGAGGCGAGAACTCGGCTGGAGTCGCGCAAGCAGATCGACGC 1259
DB |||||
1420 CGGCGGAGCTACGGCAAGCGCGAACCCTCGTACCGCTTGGCCCGCGCGCTACGACAC 1479
QY |||||
1260 CGTCTCCAGGACCGGCGTGCATGCTGACCCCGCACCGCTACGCTTGGCCAAAGAGCT 1319
DB |||||
1480 TGCCTTGACACATTCGACGCTCTGTGTGATGCAACGCTGCCCTAGCTCGCATCGAAT 1539
QY |||||
1320 GTTAGCGGTGCGCAAGACACCATGTCATGATCCACGATGCCAGGATGAGGGCAATGCCATCT 1379
DB |||||
1540 GCGCGGAGGACGATAGATCGTCAACCTTCATCAC---CAAGGCTCTCGGGATGATCGC 1596
QY |||||
1380 CACACGTCGCGCTGACCTCACCGGTCAACCGCGCTGACGTCGCCACCGGTCGCGG 1439
DB |||||
1597 CACACGCGACCATTCGACGTCGCGGACATCGTCCCTTCTCGTTCGCGCGCGCTG-- 1654
QY |||||
1440 CGAAGAGGCTGCGCTGGCTTCCAGTGTCAAGTGTAGCCGCCACTTCGAGGAGTCGACGT 1499
DB |||||
1655 -GTGAACGGCTTCCGTCGGAATGATGATCACCGGACAGACATTCGACGATGCGACGT 1713
QY |||||
1500 CTACCGCACCGGCGCGGTGATCGA 1523
DB |||||
1714 CTTCTGTGCGGACGCGCATTCGA 1737

RESULT 4
AAT62385
ID AAT62385 standard; DNA; 1566 BP.
XX
AC AAT62385;
XX
DT 24-JUN-1997 (first entry)
XX
XX Rhodococcus rhodochrous amidase coding sequence.
DE
XX Rhodococcus rhodochrous; amidase; probe; amplification; primer; PCR;
KW polymerase chain reaction; phage; open reading frame; nitrile hydratase;
XX enzyme; genetic engineering; ds.
XX Rhodococcus rhodochrous.
XX
Key Location/Qualifiers
CDS 289..291
FT /*tag= a
FT /transl_except= seq: GCA, a.a.: Asp
FT 316..318
FT /*tag= b
FT /transl_except= seq: ATG, a.a.: OTHER
FT /note= "OTHER designated Mec in specification"
XX
PN JP09009973-A.
XX
XX 14-JAN-1997.
PD
XX 27-JUN-1995; 95JP-00184934.
PF
XX 27-JUN-1995; 95JP-00184934.
PR
XX (CHCC) CHISSO CORP.
PA
XX WPI; 1997-126430/12.
DR

DR P-PSDB; AAW14159.
XX Rhodococcus rhodochrous nitrile hydratase and amidase genes - useful for
PT the industrial preparation of useful cpds.
XX
XX
PS Claim 1; Page 5-7; 16pp; Japanese.
XX
CC This is the nucleotide sequence encoding an amidase from Rhodococcus
CC rhodochrous strain IF015564. The coding sequence was isolated by
CC screening a genomic R. rhodochrous DNA library in EMBL3 vector, using a
CC probe amplified by the primers T62388-9. The primers amplified a 0.7 kb
CC fragment of the amidase gene. The library screen isolated a phage
CC containing a 15 kb insert. When analysed the insert was seen to contain
CC an approx. 6 kb SacI-ScoRI fragment (see T62387) containing the open
CC reading frames for the amidase gene, the nitrile hydratase alpha and beta
CC subunits (T62386) and another open reading frame encoding a 399 amino
CC acid protein (W14162) of unknown function. The enzymes, prepared by
CC genetic engineering methods, can be used in the industrial production of
XX useful compounds
XX
SQ Sequence 1566 BP; 318 A; 519 C; 456 G; 273 T; 0 U; 0 Other;

Query Match 11.1%; Score 175.4; DB 2; Length 1566;
Best Local Similarity 48.5%; Pred. No. 2.4e-21;
Matches 637; Conservative 0; Mismatches 661; Indels 15; Gaps 5;

Qy 217 CCGACCGGCGACGAAGACCGGTTCAACGCGATTATCCGGTTTCGCGGGTGAGGGCGCC 276
Db |||||
Qy 190 CCAACAGCGAGCGAAATCTTTGAGCGCTTGTATGTGACCAACAAGCATCCCGCGCAGC 249
Db |||||
Qy 277 ACGGAGGGGCACTGAGCGCACTTGACCGCGGATCAAGGACTGTCATGCCATCGCCGCT 336
Db |||||
Qy 250 TCGGACGGCGTCTGACCGGCGCACGCGTGGCGATCAAGGCAAACTGACCGCTGGCGGA 309
Db |||||
Qy 337 ATGCCACACGAACCGGTCGCGATGCTCCCGACTGTATGCCACGAGGATGCCGCTG 396
Db |||||
Qy 310 GTTCCGATGATGAACGGGCTCTCGGACAGTAGAGGGGTTCACTCGCTCGGACGCGACT 369
Db |||||
Qy 397 GTGGTGGAGCGGCTGTGCGCGGAGCGCCACCATCTGCGCAAGACGAACCTCGAGGAC 456
Db |||||
Qy 370 GTGATCACTGCACTACTGCGGCGCGGTGCAACCGTCTCGGGCAAGACTGTGTGAGGAC 429
Db |||||
Qy 457 ATGGCGAT---GGGTATCGGTGAAGCAGCGGTCTACGGTCTCTGCTGTAACCGCAACAC 513
Db |||||
Qy 430 CTGTGTTTCTCCGGTTTCGAGCTTTCACACCGGCAAGCGGACCGGTCCGCAATCCATGGGAC 489
Db |||||
Qy 514 CCGGCCACGCGACGCGGTGATCTTCCAGCGCTCCGGCGCTCCGCTGCGCGCATG 573
Db |||||
Qy 490 CCACGCGTGAACGAGTGGATCATTCGGTGGCAGTGGGCTCTCTCGCAACCGGTGAC 549
Db |||||
Qy 574 GTGACTTCGCTCGGCGTTCGATGAGGAGGAGGAGGATCCGGATCCCGCGCGATGGTGC 633
Db |||||
Qy 550 GTGATTTTCCATCCGCGGGATCAGGTTGGATCGATCCGGATCCCGCGGCAATTCCTGC 609
Db |||||
Qy 634 GGAAGTGTGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 693
Db |||||
Qy 610 GGGTGTGCGGACACAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 669
Db |||||
Qy 694 GACCACACCTTGGACCACTCGGCGCCCATCACAGGGGGGTTCAGGCTCAACGCGCGGGTC 753
Db |||||
Qy 670 GAGCGAACAATCGACCACTCTCGGCGCGATCACAGCGACGCTCCACGATGCGCGCACTGATG 729
Db |||||
Qy 754 CTCGAGTGTGCGCGGGCGGACTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 813
Db |||||
Qy 730 CTCTCGGTGATCGCGGGTTCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 786
Db |||||
Qy 814 CCGGAGAACTACGGCTCCGCGCTCGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 873
Db |||||
Qy 787 GCAGTGCATCTGTCTCAACCTCGACTCCGATGTGATGTGATGTGATGTGATGTGATGTG 846
Db |||||
Qy 874 GAGGAGTCACTGGAGCGGAACGGGTGCGACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 933
Db |||||
Qy 847 CGAGAGAGGTTTCGGGCGACGCGGTTCTACAGCCCGGAGGTCCAGACGCGAGTCCCGCAGCG 906
Db |||||

RESULT 5

ACA37683
ID ACA37683 standard; DNA; 1479 BP.
XX ACA37683;
AC ACA37683;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #19340.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Mycobacterium avium.
XX
PN WO20027183-A2.
XX
PD 03-OCT-2002.
XX
PP 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00915242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI: 2003-029926/02.
DR P-PSDB; ABU33813.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 25553; 1766pp; English.
XX

CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences

XX SQ Sequence 1479 BP; 199 A; 542 C; 526 G; 212 T; 0 U; 0 Other;

Query Match 11.1%; Score 175.2; DB 7; Length 1479;
Best Local Similarity 49.6%; Pred. No. 2.6e-21;
Matches 636; Conservative 0; Mismatches 613; Indels 33; Gaps 6;

QY 285 GCCACTGAGCGACCTGACCGCGCGGATCAAGGACTGTCGCCATCGCGGTATGCCAC 344
DB 213 GCCTGTGGCGGGTGCCTGGCGCTCAAGGACGCTTCCACCACGTCGACATGCCAC 272
QY 345 CACGAAACGGTCCCGATGCTCCGACTGTGATCCGACCGAGATGCGGTGGTGGGA 404
DB 273 CACTTGGGGTCCAGATTCTCAGGGGTGGGGTTCGCGGTACGACGCCACCGTCAC 332
QY 405 GCGGTGTCTCGCGGAGGCGCCACCATGTCGGCAAGACGAACCTCGAGGACATGGCGAT 464
DB 333 GAAGGTGCGGGCGCGGGCATCCCGATCTCTGGGCAAGACCAACATGGACGATTCGAT 392
QY 465 GGGTATC---GGTGAAGGAGCGGTCTACGGTCTCGGTGAAACCGAACAACCGGCCA 521
DB 393 GGGCTCTCGACCGGAGAACTCGGCTCTACGGCCCGACCGCGCAACCGTGGATGTCGAG 452
QY 522 CGGCACGCGGTGATCTTCAGCGGGTCTCGCGCGCTCGGTGCGCGCATGGTCGACTT 581
DB 453 GTTGGCCGCGGTTCGCGGGCGGCGAGCGCGCGCTGGCGGTTCAGGCGCGCGCT 512
QY 582 CGCCCTGGCGTCTGATGAGGAGGAGAGATTCGGATTCGCGCGCGCATGGTTCGAGCTGGT 641
DB 513 GCGCATCGGCTCCGACACCGCGGGTTCGATTCGGGAGCGCGCGCGCTGACCGGACCGT 572
QY 642 CGGCATGAAGGCGACCCACCGGCTGGTGGCGCTTACGGCTGACATACATGGACCCAC 701

DB 573 CGGGGTCAAGCCACCTACGGCACGGTGTCCCGGTACCGGCTGGTGGCTTCGCGTCAATC 632
QY 702 CTTTGGACCAATCATCGGCGCCATCACAGGGGGGTTCGAGCTCAACGCGCGGGTCTCTCAGGT 761
DB 633 GCTGGATCAGGGCGGCGCGTGGCGCGCACCGTGTTCGACACCGCGCTCTGCACAGGT 692
QY 762 GTTGGCGGGCGGACTGGCGGACCTCAGTGGGTGCGTAACCTTCCGAGGCGGAGAA 821
DB 693 GATCGCGGCGCACGATCCGCGACTCCACCTCCGT---CGACGCGCGCGTGGCGCGACGT 749
QY 822 CTACGGTCTCGCGCTCGGGGAGGAGATTCGGTCTTGAGATTCCGGTCTCGTTCGAGGAGTC 881
DB 750 CGTCCGCGCGCGCGCGCGCGCGCGCGCTGACCTCAAGGTGTTCGGTTCGGGTGGT 809
QY 882 ACTGAGAGCGCAAGCGTGGG-----ACGCGGACGCTGATCGCGGCTTCAACAGGAGT 935
DB 810 CAAGCAGCTGCGGGCGGAGGATATCAGCCCGCGGTGCTGGGCTCTCTCGAGCGCGCGT 869
QY 936 GCGCGCGCT---CGAGAGCGCGGTCGACACCTCGAGCGGGTCTCGGTGCCGTTCGAGC 992
DB 870 CGAGCAGTTGACCGCGCTGGCGCGCGAGGTACGAGGTTCGACTGCGCGCATCTTCGAGTA 929
QY 993 GCGCGCTGGCTTATCAGAGCGCGCTGATGGCTTTCAAGCGCGCGCTATGCGGAGTTC 1052
DB 930 CGCGCTGGCGCGCTACTACTGATTTTGGCTTCGAGGTGTCCAGCAACCTTCGCGCGCT 989
QY 1053 CGCGCGTGGGGTACTTCCACAAGGGCGCGTGGAGCTCAGCACCGCGCTCACAGCGG 1112
DB 990 CGAGCGATGCGCTACGGGTCGCGCATCGGCGAGCGAGCGAGCGAGCGAGGT 1049
QY 1113 GCGCCAGAGTGGCACCCACCAAGGATCTGGGATCTGTCCCGGCTGATGCTGGTGTAT 1172
DB 1050 GATGGCGCTGACCCGCGCGCGCGGTTCGCGCGCGAGGTCAAGCGCGCATCATGATCG 1109
QY 1173 CGCGGAGACCTCGCGGACGAAATACCTCGGCATCCACTACGCGAGGCGCAGAACTCGC 1232
DB 1110 CACTAGCGCTGTCCGCGGCTACTACGAGCGCTACTAACACCAGGCGCGAGGTGG 1169
QY 1233 GCTGAGCTGGGCAAGCAGATCGACCGCTCTTCAGGACCGGGTGTGACTGCTGACCC 1292
DB 1170 CACCTGATGCGCGCGGATCTCGACCGCGCTACGAGTCCGTCGAGTGGTGGTCTCG 1229
QY 1293 GACACGCTTACCGTTGCCAAGCGCTGTGAGCGGTCCGCAAGCACCATGTCCATGAT 1352
DB 1230 GGCACCCCGACACCGCGCTTCGGGTGGGGGAGAGGTTCGACGACCCCGCTGGCCATGA 1289
QY 1353 CCACCGGATGACGGGCAATGCGATCTCAACAGCTGCCCGCTGGACCTCACCGGTCA 1412
DB 1290 TCTGTTGCA-----CCTGTGACGCTGCGCTGAACCTGGCGGGCACTG 1334
QY 1413 GCGCTGACGGTCCACCGGTG---CGGGCGAGAGGCGCTTCGCGGTGGCTCCAACT 1469
DB 1335 CGGCATGTGGTGGCATCCGGATTTGTCGCCCGCAGACGACCTGCGCGGTGGGTTCAGAT 1394
QY 1470 GATAGCGCGCACTTCGAGGAGTTCGACGCTCTACCGCACCGCGCGCGCTGATCGAGCGCG 1529
DB 1395 CATGGCGCGCGCTGCGCGACCGGTGTGACCGGGTTCGCGCGCGCTACGAGGCGCG 1454
QY 1530 CGGCTATGGGAGCTCGCGCG 1551
DB 1455 GCGCGCGCGCTGCGAGCGCG 1476

RESULT 6
AAI99683_33/c
Continuation (34 of 44) of AAI99683 from base 3300001 (Mycobacterium tuberculosis strain:
WP Sequence split into 44 fragments LOCUS AAI99683 Accession AAI99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000

WP	AA199683_04	400001	510000	692	TGACCAACCTTGGACCATCGGGCCCATCACCAGGGGGTTCAGCTCAACGCCCGGG	751
WP	AA199683_05	500001	610000	65141	CGCGTCTCTGCTGATCAGGCGCCCGCTGTGCGCGACCGTCTTGGACACCCGCTGT	65082
WP	AA199683_06	600001	710000			
WP	AA199683_07	700001	810000			
WP	AA199683_08	800001	910000	752	TCCTCAGAGTGTGGCCGGGCGGACTGGCGCGACCTTCAGTGGGTGCTAAACCTTCGG	811
WP	AA199683_09	900001	1010000	65081	TGCATCAGGTGATCGCGGCGCACGACCGCGCGACTCCAGTCGCTCGAGCCGAGTGC	65022
WP	AA199683_10	1000001	1110000			
WP	AA199683_11	1100001	1210000			
WP	AA199683_12	1200001	1310000	812	AGCCGGAGAACTACGGCTCCGGCTCGCGGAGGAGATCCGGTCTGAGATTCGGGTTCG	871
WP	AA199683_13	1300001	1410000			
WP	AA199683_14	1400001	1510000	65021	CCGACGTGTGGCGCCGCTAGGCGCCGCGCGGTTCGGGGATCTCGTGGGTGCGGGTTCG	64962
WP	AA199683_15	1500001	1610000			
WP	AA199683_16	1600001	1710000			
WP	AA199683_17	1700001	1810000	872	TCGAGGAGTCACTGGAGCCGAACGCTGCG-----ACGCGGAGAGTGTGATCGCCCGTTCA	925
WP	AA199683_18	1800001	1910000			
WP	AA199683_19	1900001	2010000	64961	CGTGGTTTCGACAGCTGCACGGCGCGAGGGCTACACAGCCGGGCTGCTG3CCTCTTCG	64902
WP	AA199683_20	2000001	2110000			
WP	AA199683_21	2100001	2210000	926	ACGAGGACTGGCGCGCTCGAGAGCCCGGTGCGACCATCGAGCGGTCTCGGTGCGGT	985
WP	AA199683_22	2200001	2310000			
WP	AA199683_23	2300001	2410000	64901	AGGCTGCGGTGGAGCAGCTAACCGCGTGGCGGTGAGGTGAGCGAGGTGAGTCCCGCGC	64842
WP	AA199683_24	2400001	2510000			
WP	AA199683_25	2500001	2610000	986	TGTGAGCGGCGCTGGCTATCCAGAGCGCGTGTGGTCTTCAACG---CGCGCGCTA	1042
WP	AA199683_26	2600001	2710000			
WP	AA199683_27	2700001	2810000	64841	ACTTCGACCATGCCCTGGCGGCTATTACCTGATCTTGCTCGGAGGTCTCGAGCAATC	64782
WP	AA199683_28	2800001	2910000			
WP	AA199683_29	2900001	3010000			
WP	AA199683_30	3000001	3110000	1043	TGGCGAGTCCCGCGGTGTGGGTACTTTCACAAAGGGGGCGGTGAGAGTCAAGCAACCGCG	1102
WP	AA199683_31	3100001	3210000			
WP	AA199683_32	3200001	3310000	64781	TGGCGCGCTTCGACGCGATCGCTACGCGGTGCGGCTCGGCGACGACGCGACCGCGAGC	64722
WP	AA199683_33	3300001	3410000			
WP	AA199683_34	3400001	3510000	1103	TCAGACGCGGCGCCAGAGTCGACACCAACCAAGAGTCTGGCGATCTGTCCCGGTGA	1162
WP	AA199683_35	3500001	3610000			
WP	AA199683_36	3600001	3710000	64721	CCGAGGAGTGTATGCGCATGACCCGCGCGCGGTTCGGGCGCGAGGTCAAGCGGCGCA	64662
WP	AA199683_37	3700001	3810000			
WP	AA199683_38	3800001	3910000	1163	TGCTGGTGTATCGGAGGACCTGCGCGACGAATACCTCGGATCCACTACGCGAAGCGC	1222
WP	AA199683_39	3900001	4010000			
WP	AA199683_40	4000001	4110000	64661	TCATGATCGGCACCTAGCGTGTGCGCGGCTACTACGACGCTATTACAACCGCGCGC	64602
WP	AA199683_41	4100001	4210000			
WP	AA199683_42	4200001	4310000	1223	AGAACTGCGGCTGAGCTGGCAAGCAGATCGAGCGCGCTCTCCAGAGCCGGCTGCAC	1282
WP	AA199683_43	4300001	4400001			
WP	AA199683_44	4400001	4500001	64601	AGAAAGTGGCAGCAGCTGATCCCGCGGACCTCGAGCGGCGGTATCGGTCCGTGACGTGC	64542
WP	AA199683_45	4500001	4600001			
WP	AA199683_46	4600001	4700001	1283	TGCTGACCCGACCAACCTACCTGCTTCCCAACGAGCTGTGAGCGGTGCGCAAGACCA	1342
WP	AA199683_47	4700001	4800001			
WP	AA199683_48	4800001	4900001	64541	TGGTGTGCGCCACGACCCGACCCGCTTCCGGTGGGTGAGAGGTGAGACGATCCGC	64482
WP	AA199683_49	4900001	5000001			
WP	AA199683_50	5000001	5100001	1343	TGTCATGATCCCAAGGATGACGGGCAATGGGATCCCTCAACAGTGCCTGCGCTGACCTCA	1402
WP	AA199683_51	5100001	5200001			
WP	AA199683_52	5200001	5300001	64481	TGGCGATGTACTTGTTCGA-----CCTGTGACGCTGCCCTGAACTTGG	64437
WP	AA199683_53	5300001	5400001			
WP	AA199683_54	5400001	5500001	1403	CCGCTCACCCGCGCTGACGCTGCCACCGGTG---CGGGCGAGAGGGGCTGCCCGTGTG	1459
WP	AA199683_55	5500001	5600001			
WP	AA199683_56	5600001	5700001	64436	CCGCGCACTGCGGCAATGCTGTGCGCTCGGGCTCTCCCGGAGCAGCGGTTCGCGGTG	64377
WP	AA199683_57	5700001	5800001			
WP	AA199683_58	5800001	5900001	1460	GCCTCCAGTGTATAGCGCGGCACTTCAGAGAGTGCAGCTCTACCGCACCGCGCGCGTGA	1519
WP	AA199683_59	5900001	6000001			
WP	AA199683_60	6000001	6100001	64376	GCCTACAGATCATGCGCGCGGCAATTGGCCGACCGGCTTACCGGGTGGGGCGGCTT	64317
WP	AA199683_61	6100001	6200001			
WP	AA199683_62	6200001	6300001	1520	TCGAGGCGCGCGCGC	1533
WP	AA199683_63	6300001	6400001			
WP	AA199683_64	6400001	6500001	64316	ATGAGGCGCGCGCGC	64303

RESULT 7

ACA38250

ID ACA38250 standard; DNA; 1482 BP.

XX ACA38250;

XX AC

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #19907.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

KW

Query Match 11.0%; Score 172.8; DB 4; Length 110000;

Best Local Similarity 48.9%; Pred. No. 5.6e-21;

Matches 623; Conservative 2; Mismatches 619; Indels 30; Gaps 5;

QY	275	CCACGAGGGCCACTGAGCGACTGACCGCGCGATCAAGGACTGCATCGCGCATCGCG	334
Db	65561	CTTGGCTGCGCGCTGGCGGGGTGCGGTGCGGCTCAAGGACGTCTTCAACACGAGCG	65502
QY	335	GTATGCCACACGAAACGGGTCCGGATGCTCCGACTGTGATCGCACCGAGGATGCGG	394
Db	65501	ACATGCCACCACTCGGGTCAAAATCTGGAGGGATGCGATCTCCCTACGACGCA	65442
QY	395	TGCTGTGGAGCGGCTGCTCGGGGAGGCGCCACCATCTCGGCAAGAGCAACCTCGAGG	454
Db	65441	CGCTGACCGCGGGTGTGCGCGCGCGGGGATCCCGATCTCGGCAAGACCAACATGAGC	65382
QY	455	ACATGCGCATGGG---TATCGGTGAAGGAGCAGCTCTACGCTCTGCGGTGAACCCGAA	511
Db	65381	AGTTCGATGGCTGCTGACGAGAACTCCGCTTACGCTCCACCCGCAACCGGTGA	65322
QY	512	ACCCGCGCCACGCGAGGTGATCTTCAGCGGCTCCGGCGTGGCGTGTGCGGCA	571
Db	65321	ATCTGACCGGTACCCGCGGTTCCGGTGGCGGAGCGCGCGGCTGCGCGGTTCC	65262
QY	572	TGCTGACATTCGCTTGGGCGTGTGAGGAGGAGCATCGGATCCCGCGCGCATGTG	631
Db	65261	AGCGCGCTGGCATCGATYCGACACCGGGGGTCTGATCCCGACGCGCGCGCTGA	65202
QY	632	GCGGACTGTGCGCATGAGGAGCCACCGGCTGTGCGGTCTTACGCGCTGACATACA	691
Db	65201	CCGCGACCGTGGCGTCAAAACCCACCTACGCGACCGGTGTGCGGTATGGGTGGTGGCT	65142

QY 1460 GCTCAAGTATAGAGCCGCCACTTGGAGAGTCGACGCTCTACCGCAGCGCGCGTGA 1519
 Db 1388 GCCTACAGATCATGGCGCGGCGATTGGCCGACGACCGGCTCTACCGGGTGGGGCGGCTT 1447
 QY 1520 TCGAGCGCGCGCG 1533
 Db 1448 ATGAGCGCGCGCGC 1461

RESULT 8
 ABZ71125
 ID ABZ71125 standard; DNA; 1482 BP.
 XX AC ABZ71125;
 XX DT 28-APR-2003 (first entry)
 XX DE Mycobacterium tuberculosis protein encoding DNA SEQ ID NO:128.
 XX KW Mycobacterium tuberculosis; infection; antibacterial; tuberculostatic;
 KW KW immunostimulant; vaccine; gene therapy; mycobacterial infection; gene;
 KW ds.
 XX OS Mycobacterium tuberculosis.
 XX PN WO2003000721-A2.
 XX PD 03-JAN-2003.
 XX PF 21-JUN-2002; 2002WO-GB002845.
 XX PR 22-JUN-2001; 2001GB-00015365.
 XX PR 07-SEP-2001; 2001GB-00021780.
 XX PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
 XX PI James BW, Bacon J, Marsh P;
 XX P-PSDB; ABP57499.
 XX WPI; 2003-201403/19.
 XX PT New mycobacterial peptide, its fragment, variant or derivative, useful as
 PT vaccine for treating or preventing mycobacterial infections, and as
 PT diagnostic reagents for identifying such infections.
 XX PS Claim 15; Page 233-234; 246pp; English.
 XX CC ABP57436 to ABP57504 represent mycobacterial amino acid sequences (I)
 CC encoded by ABZ71062 to ABZ71130 (II), which are isolated from
 CC Mycobacterium tuberculosis. (I) are encoded by genes (II) whose
 CC expression is induced or up-regulated during culture of a mycobacterium
 CC under conditions defined by a dissolved oxygen tension of at least 10%
 CC air saturation measured at 37 plus degrees Celsius, when compared with a
 CC dissolved oxygen tension of at least 40% air saturation measured at 37
 CC plus degrees Celsius. (I) and (II) have antibacterial, tuberculostatic
 CC and immunostimulant activities, and can be used in vaccines and gene
 CC therapy. (I) and (II) can be used for the manufacture of a medicament for
 CC treating or preventing a mycobacterial infection. They can also be used
 CC for the manufacture of a diagnostic reagent for identifying a
 CC mycobacterial infection

SQ Sequence 1482 BP; 212 A; 509 C; 529 G; 232 T; 0 U; 0 Other;
 Query Match 10.8%; Score 170.4; DB 7; Length 1482;
 Best Local Similarity 48.9%; Pred. No. 1.7e-20;
 Matches 623; Conservative 0; Mismatches 621; Indels 30; Gaps 5;

QY 275 CCACGGAGGGCGCACTGACGACCTGACCGCGCGGATCAAGGACTGCGCATGCCATCGCG 334
 Db 203 CTTTGCCTCGCGCTGCGCGGGTGCCTGCGGCTCAGGAGCTCTTACCACACGCG 262
 QY 335 GTATGCCCAACACGAAACGGGTCCTCGGATGCTCCCGACTGTGATCGCCACGAGGATGCCG 394

Db 263 ACATGCCCAACACCTGCGGGTCAAAAATCTTGAGGGATGCGCATCTCCCTTCAGACGCCA 322
 QY 395 TGTGTGTGAGAGCGCTCTCGCGGAGGCGCCACCATCTGTCGCAAGACGAACTTCGAGG 454
 Db 323 CGCTGACCGCGCGTTCGCGCGCGCGGGATCCGATCTCTGCGGCAAGACCAATGAGACG 382
 QY 455 ACATGCGGATGGG---TATCGGTGAAGGAGCGTCTACGCTCTCTGCGCTGAACCCGGAACA 511
 Db 383 AGTTTCGGATGGGCTGCTCGACGAGAACTCCGCTTACGTTCCACCCGCAACCCGTGGA 442
 QY 512 ACCCGCGCCACGCGACGCGGTGGATCTTCCAGCGGCTCCGCGCTGCTGCTGCTGCTGCTG 571
 Db 443 ATCTCGACCGGGTACCCCGCGGTTCGCGTGGCGGCGCGCGCGCTGCGCGCTTCC 502
 QY 572 TGTTCGACTTTCGCGCTGCGCTGATGAGGAGGAGCAGCATCCGATCCCGCGCGCATGCT 631
 Db 503 AGGCGCGCTGCGCATCGGATCCGACACCGGCGGGTGCATCCGCGACCGCGCGCGCTGA 562
 QY 632 CGGACTGCTCGCATGAAGCGGACCCACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 691
 Db 563 CCGCGACGCTGCGCGTCAAAACCCACCTACGCGACGGTGTGCGCTATGCGCTGCTGCGCT 622
 QY 692 TGGACCAACCTTGGACCAATCGGGGCCATCACCAGGGGGTTCGAGCTCAACCGCGCGG 751
 Db 623 CGCGTCTCTGCTGGATCAGGGCGCGCGCTGTGCGCGACCGCTCTTGGACACCGCGCTGT 682
 QY 752 TCCTCGAGGTGTTGCGCGGGCGCGACTGCGGCGGACCTCACTGCGGTGCTGCTGCTGCTGCT 811
 Db 683 TGCATCAGTGTATCG 742
 QY 812 AGCGGAGAACTACGCTTCGCGCTCGCGGAGGAGATTCGCGTCTGAGATTCGCGGTGCG 871
 Db 743 CCGACGTGCTGCGCGCGCTAGGGCGCGCGCTGCGGCGATCTCGTGTGCGGTGCGGGTTCG 802
 QY 872 TCGAGGAGTCACTGAGCGCGGAAACGCTGCG-----ACGCCGAGCGTGTGATCGCGCGGTTC 925
 Db 803 CGGTGTTGCAAGCTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 862
 QY 926 ACCAGGAGTTCGCGCGCTCGAGAGCGCGGTGCGGACCATCGAGCGGGTCTCGGTGCGGT 985
 Db 863 AGGCTCGCGTGGAGCAGCTAACCGCGCTGGCGCGCTGAGGCTCAGCGAGGTTCGACATTC 922
 QY 986 TGTGAGCGCGCGCTGCGCTTATCCAGAGCGCGGTGATGGCTTTCAACG---CGCGCGCTA 1042
 Db 923 ACTTCGACCATGCTTCGCGCGCTTATCTGATCTTCTGCTCGGAGGTTCGAGCAATC 982
 QY 1043 TGGCGGACTTCGCGCGGTGTGGGCTACTTCCAAAGGGCGCGGTGAGACGTACGACCGCG 1102
 Db 983 TGGCGCGCTTCGACCGGATGCGCTACGGCTGCGGGTTCGCGGACGACGCGCACCCGAGCG 1042
 QY 1103 TCACAGCGCGCGCGCGAGTTCGACCAACCCACAGGATCTGCGGATCTGCTGCGCGCTGA 1162
 Db 1043 CCGAGGAGGTGATGCGGATGACCCGCGCGCGCGGTTCGCGCGCGCGGATCAGCGCGCGCA 1102
 QY 1163 TGCTGTGATCGCGGAGCACCTGCGCGACGMAATACCTCGGATTCACCTACGCGAAGGCGC 1222
 Db 1103 TCATGATCGGACCTACCGCTTGTGCGCGCGCTACTACGACGCTATTACACACGAGGCG 1162
 QY 1223 AGAACCTGCGGTGAGCTCGCGAGGAGATCGAGCGCGCTCTCCAGGACCGCGGTGCGAC 1282
 Db 1163 AGAAGGTGCGCACGCTGATCGCGCGCGCGCTTCGCGCGCGGTATCGGTTCGTCGACGTGC 1222
 QY 1283 TGTGACCGCGCGCGCGCGCTTACCGCTTTCGCAACGAGCTGTTGAGCGGTTCGCGCAAGACCA 1342
 Db 1223 TGGTGTGCG 1282
 QY 1343 TGTTCATGATCCACGAGATGAGCGGGAATGCGATCTCTMAACAGCTGCGCGCTGCGCTCA 1402
 Db 1283 TGGCGATGACTTGTTCGA-----CCTGTGCACGCTGCGCGCTGCGCTGCGCTGCG 1327
 QY 1403 CGGTCACCGCGCGCTGAGCGGTGCCACCGG---TCCGCGCGCGAGGCGCTTCGCCCGTTG 1459

Db 1328 CCGGCCACTGCGGCATGCTGTGCGCTCGGGCTCTCCCGGACGACGGTTGCGGGTTG 1387
Qy 1460 GCTTCAAGTATAGCCGCCACCTTCGAGGAGTCAGCGCTACCGCACCGCGCGCGTGA 1519
Db 1388 GCTTCAAGTATAGCGCGCGCATTTGGCGGACGACCGGCTTACCGGTGGGGCGGCTT 1447
Qy 1520 TCGAGGCGCGCGG 1533
Db 1448 ATGAGGCGCGCGC 1461

RESULT 9

ACA40772

ID ACA40772 standard; DNA; 1485 BP.

AC ACA40772;

XX 19-JUN-2003 (first entry)

DT 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #22429.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

KW drug design; Gene.

XX Mycobacterium tuberculosis.

OS

XX W020027183-A2.

PN 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX P-PSDB; ABU36902.

XX New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 28642; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of

XX CC the 6213 antisense sequences given in the specification where expression

XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:

XX CC (1) a vector comprising a promoter operably linked to the nucleic acid

XX CC encoding a polypeptide whose expression is inhibited by the antisense

XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

XX CC polypeptide or its fragment whose expression is inhibited by the

XX CC antisense nucleic acid; (4) an antibody capable of specifically binding

XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

XX CC proliferation or the activity of a gene in an operon required for

XX CC proliferation; (7) identifying a compound that influences the activity of

XX CC the gene product or that has an activity against a biological pathway

XX CC required for proliferation, or that inhibits cellular proliferation; (8)

XX CC identifying a gene required for cellular proliferation or the biological

XX CC pathway in which a proliferation-required gene or its gene product lies

XX CC or a gene on which the test compound that inhibits proliferation of an

XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

XX CC compound's activity; (11) a culture comprising strains in which the gene

XX CC product is overexpressed or underexpressed; (12) determining the extent

XX CC to which each of the strains is present in a culture or collection of

XX CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1485 BP; 213 A; 509 C; 530 G; 233 T; 0 U; 0 Other;

SQ Query Match 10.8%; Score 170.4; DB 7; Length 1485;

Best Local Similarity 48.9%; Pred. No. 1,7e-20;

Matches 623; Conservative 0; Mismatches 621; Indels 30; Gaps 5;

Qy 275 CCACGAGGGGCCACTGAGCGACCTGACCGCGGATCAAGGACTCATCGGCATCGCG 334

Db 203 CCTTGGCGTGGCGCTGGCGGGTGGCGTCAAGGACGCTTCAACCCAGCG 262

Qy 335 GTATGCCACACGAGAAAGGGTCCCGGATGCTCCCGACTGTGATCGCCACCGAGGATGCCG 394

Db 263 ACATGCCACACCTGGGGTCAAAAATCTTGAGGGATGGCGATCTCCTACGACGCCA 322

Qy 395 TGGTGTGAGAGCGGCTGCTCGCGGAGGGCGGACCATCTGCGGCAAGAGAACTCGAGG 454

Db 323 CGTGAACCGCGCGGTTGCGCGCGCGCGGATCCCGATCTTGGGCAAGACCAATGAGCG 382

Qy 455 ACATGGCGATGGG---TATCGGTGAAGGAGCGGCTCTACGGTCTCGGCTGAAACCCGAACA 511

Db 383 AGTTGGGATGGGCTCGTCAACGAGAACTCCGCTTACGGTCCACCCGACCCGCTGA 442

Qy 512 ACCCGGCCACGCGACGGGTGGATCTTCCAGCGGCTCGCGCGTGGCGCTGCGTCCGCGCA 571

Db 443 ATCTCGACCGGTACCGCGGCTTCCGGTGGCGGAGCGCGCGGCGTGGCGCGCTTC 502

Qy 572 TGTCTGACTTCCGCGCGGCGTGGATGAGGAGGAGCATCCGATCCCGCGCGCATGGT 631

Db 503 AGGCGCGCTGGCGATCGGATCCGACACCGGGGGTGGATCCCGAGCGCGCGCTGA 562

Qy 632 GCGGACTGTGCGGATGAAGCGACCGACCGCTGTGCGGCTTTACGGCTGACATACA 691

Db 563 CCGGACCGTGGCGGTCAAAACCCACCTACGCGACGGTGTGCGCTATGGGCTGGGCT 622

Qy 692 TGGACACACATTGGGACCATCGGGCCATCAACAGGGGGTGGAGTCAACCGCGCGG 751

Db 623 GCGGCTCTCGTGGATCAGGGCGGCGCGTGTGCGGCGACCGTCTTGGACACCGCGTGT 682

Qy 752 TCCTCGAGGTGTGGCGGGGCGGACTGGCGGACCGCTCAGTGGTGGTAACTTCCGG 811

Db 683 TGCAATCAGGTGATCGCGGCGCACGACCGCGCGACTCCAGTCTGGTCGACGCGAGGTGC 742

Qy 812 AGCGGAGAACTAGGCTCCGCGCTCGGCGAGGAGTATCCGGTCTGAGATTCCGCGTGC 871

Db 743 CCGACGTGTGGGCGCGCTAGGGCGCGCGCTGGGGGATCTGCGTGGCGTGGGCTGC 802

Qy 872 TCAGGAGTCACTGGAGCCGAAACGGTGC-----ACGCGGACGAGTATCGCGCGTCA 925

Db 803 GCGTGGTTCACAGCTGACGCGCGGAGGGCTACCAAGCGCGCGTGTGGCTCTCTTCG 862

Qy 926 ACAGGAGCTGGCGCGCGCTCGAGCGCGCGTGGACCATCGAGCGGGTCTCGGTCCGT 985

Db 863 AGGCTGCGGTGGAGCAGCTAACCGCGCTGGCGCGCTGAGGTCAGCGAGGTGAGTGC 922

Qy 986 TGTGAGCGCGCGCTGGCGCTTATCCAGAGCGCGCTGATGGCTTTCAAG---CGCGGCTA 1042

Db 923 ACTTCGACCATGCGCTGGCGCGCTATTACCTGATTTCTGCTCGAGGTGTCGAGCAATC 982

Qy 1043 TGGCGGACTCGCGGTGTGGGCTACTTCCACAAGGGCGCGTGGAGTCAAGCACCGCG 1102

Db 983 TGGCGCGCTTCAGCGCGATCGCTACGGGTGGCGGTGGCGGACGAGCGGACCGCGAGCG 1042

QY	1103	TCACGAGCGCGCCAGAGTGCACACCCACAGGATCTGGCGATCTGTCCCGGTGA	1162
Db	1043	CCGAGAGGTGATGGCGATGACCCGGCGCGGTTTCGGGCCCGAGGTCAGCGCGCA	1102
QY	1163	TGCTGTGATCCGGAGACCTTGGCGACGAATACCTCGGCATCCACTACGGGAAGGCG	1222
Db	1103	TCATGATCGCACCTACGCGTTGTGGCGGCTACTACGACGCTTATTACAAACGAGGCG	1162
QY	1223	AGACTCGCGGTGGAGCTCGCAAGCAGATCGACGCGTCTCTCCAGAGCGGGCTGCAC	1282
Db	1163	AGAGGTGCGACGTGATCGCCGGGACCTCGACGGCGTATCGTCCGTGCGACGTGC	1222
QY	1283	TGCTGACCCCGACCGCTACCGTTGCGCAAGAGCTGTGACGGTTCGGCAAGACCA	1342
Db	1223	TGCTGTGCGCCACGACCCCGACCGCGTTTCGGATGGGTGAGAAGTGGACGATCCGC	1282
QY	1343	TGTCATGATCCACCGATGACGGGCAATGCCATCTCAACAGCTGCCGCTGGACCTCA	1402
Db	1283	TGCGGATGATCTTGTTCGA-----CCTGTGACGCTGCGCTGAACTTGG	1327
QY	1403	CGGTCAACCGCGCTGACGGTCCCGACCGG---TGCGGGCGAAGAGGCTTSCCGTTG	1459
Db	1328	CGGCGCACTGGCGCATGTCTGTGCGCTCGGGCTCTCCCGGACGACGGTTTGCCTTG	1387
QY	1460	GCTTCAAGTATAGCGCGCACTTCGAGAGTTCGACGCTCTACCGCACCGCGCGGTGA	1519
Db	1388	GCCTACAGATCATGGCGCGCGCATTCGCCGACGACCGGCTCTACCGGTTGGGGCGGCTT	1447
QY	1520	TCGAGCGCGCGCG 1533	
Db	1448	ATGAGCGCGCGCG 1461	

RESULT 10
AAI99682_33/c
Continuation (34 of 45) of AAI99682 from base 3300001 (Mycobacterium tuberculosis strain
WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	210001	310000
WP	AAI99682_03	310001	410000
WP	AAI99682_04	410001	510000
WP	AAI99682_05	510001	610000
WP	AAI99682_06	610001	710000
WP	AAI99682_07	710001	810000
WP	AAI99682_08	810001	910000
WP	AAI99682_09	910001	1010000
WP	AAI99682_10	1010001	1110000
WP	AAI99682_11	1110001	1210000
WP	AAI99682_12	1210001	1310000
WP	AAI99682_13	1310001	1410000
WP	AAI99682_14	1410001	1510000
WP	AAI99682_15	1510001	1610000
WP	AAI99682_16	1610001	1710000
WP	AAI99682_17	1710001	1810000
WP	AAI99682_18	1810001	1910000
WP	AAI99682_19	1910001	2010000
WP	AAI99682_20	2010001	2110000
WP	AAI99682_21	2110001	2210000
WP	AAI99682_22	2210001	2310000
WP	AAI99682_23	2310001	2410000
WP	AAI99682_24	2410001	2510000
WP	AAI99682_25	2510001	2610000
WP	AAI99682_26	2610001	2710000
WP	AAI99682_27	2710001	2810000
WP	AAI99682_28	2810001	2910000
WP	AAI99682_29	2910001	3010000
WP	AAI99682_30	3010001	3110000
WP	AAI99682_31	3110001	3210000
WP	AAI99682_32	3210001	3310000
WP	AAI99682_33	3310001	3410000
WP	AAI99682_34	3410001	3510000

WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529

Query Match 10.88; Score 170.4; DB 4; Length 110000;
Best Local Similarity 48.94; Pred. No. 1.5e-20;
Matches 623; Conservative 0; Mismatches 621; Indels 30; Gaps 5;

QY	275	CCACGAGGGGCGCACTGAGCGACCTGACCGCCCGCGATCAAGGACTGCATCGCATCGCG	334
Db	71229	CCTTGCGCTCGGCGCTGGCGGGTGCGGTGCGCTCAAGGACGCTTCCACCCAGCG	71170
QY	335	GTATGCCACACGAAACGGGTCCCGGATGCTCCGACTGTGATCGCCACCGAGATCGCG	394
Db	71169	ACATGCCACACCGCTGCGGGTCAAAAATCCTGGAGGATGGCGATCTCCCTACGACCGCA	71110
QY	395	TGCTGTGGAGCGCTGCTCGCGCAGCGCGCCACATCGTCGCGCAAGCAACCTCGAGG	454
Db	71109	CGCTGACCGCGCGTTGGCGCGCGGGGATCCCGATCTCTGGGCAAGCAACATGACG	71050
QY	455	ACATGGCGATGGG---TATCGGTGAAGCGACGCTCTACGCTCTCGCTGACCGCAACA	511
Db	71049	AGTTTCGCGATGGGCTCGTCCGACGGAGAACTCCGCTTACGCTCCACCGCAACCGGTGA	70990
QY	512	ACCCGCGCCACGCGCGGTGGATCTTCCAGCGGCTCCGCGCTGCGCTGCTGCCGCGCA	571
Db	70989	ATCTGACCGGTACCCGCGGTTCCGTTGGCGGCGCGCGCGCGTGGCGCGGTTC	70930
QY	572	TGCTGACTTCGCTCGGCTCGATGAGGACGACGATCCGATCCGATCCGCGCCGATGCT	631
Db	70929	AGGCGCGCTGCCATCGATCCGACACCGCGGGGTGATCCGCGACGCGCGCGCTGA	70870
QY	632	GCGGACTGTCGGATGAAGCGCACCGCGCTGCTGCGCTCTTACGCGCTGCACATACA	691
Db	70869	CCGCGACCGTCCGCGCTCAACCCACCTACCGCACGGTGTGCGCTATGGGTGCTGG	70810
QY	692	TGGACACACCTTGACACACATCGGGGCCATCAACGAGGGGGTGGAGTCAACGCGCGG	751
Db	70809	GGCGTCTCTGCTGGATCAGGGCGCGCGCTGCGCGCACCGCTCTTGGACACCGCGTGT	70750
QY	752	TCCTCAGGTGTGCGCGCGCGCGCTGCGCGCGCGCTCAGTGGGTGCGTAACCTTCGCG	811
Db	70749	TGCATCAGGTGATCGCGCGCGCGCGCGCGCTCCACGTCCGCTCGACCGCGAGTGC	70690
QY	812	AGCCCGGAGAACTACCGCTCCGCGCTCGCGCGAGGAGTATCCGCTGTGAGTTCGCGGTG	871
Db	70689	CCGACGTGTGGCGCGCGCTAGGGCGCGCGCGCTGCGGGATCTTGGTGGCGTGGCGTGC	70630
QY	872	TCGAGAGTCACTGAGCGCGCGCGCGCGCGCGCTGCGCGCGCGCTTCCCGCGGTCA	925
Db	70629	GCGTGTTCGACAGCTGACCGCGCGCGCGCGCGCTACGACCGCGCGCGCTTGGCGCTCTT	70570
QY	926	ACCAGGACTGCGCGCGCTCGAGAGCGCGCGCTGCGCGCGCTCGAGCGGCTCTCGGTGCG	985
Db	70569	AGGCTGCCGTGAGCAGCTAACCGCGCTGGCGCTGAGGTGAGCGAGGTGCGACTGCCCG	70510
QY	986	TGTGACGCGCGCTGGCGCTATCCAGAGCGCGGTGATGGCTTTCAAAG---CGCGGCTA	1042
Db	70509	ACTTCGACCATGCCCTGGCGCGCTATTACTGCTGATCTGCCCTCGAGGTTGCGAACAATC	70450
QY	1043	TGGCGGACTCCCGCGGTGGGCTACTTCCACAAGGGCGCGGTGGACGTCAGCACCGCGCG	1102
Db	70449	TGGCGGCTTTCACGCGATCGCTACGGGCTCGCGGTGCGCGACGACCGCGCGCGCG	70390
QY	1103	TCACGAGCGCGCGCGCGAGTTCGCAACACCAAGGATCTGTCGCGATCTTGTCCCGGTGA	1162

DB 70389 CCGAGGAGTGATGCGATGACCCCGGCGCGCGTTTCGGGCCCGAGGTCAAGCGGCGCA 70330
QY 1163 TGCTGGTGTATCGCGGAGCACCTTCGCGACGAATACCTCGGCATCCACTACGCGAAGGCGC 1222
DB 70329 TCATGATCGGACCTACGGGTGTGCGCGGCTACTACGAGCCTATTACACAGGCGC 70270
QY 1223 AGAACCTCGGGTGGAGTTCGCGAAGAGATGACGCGCTCTCCAGAACCGGGTGTGAC 1282
DB 70269 AGAAGGTGCGACGTGTATCGCGCGCGACCTCGACGCGCGGTATCGGTCCGTCGACGTGC 70210
QY 1283 TGCTGACCCCGACCGCTACCGTTGCAAGAGCTGTGAGCGGTGCGCAAGACACCA 1342
DB 70209 TGGTGTGCGGACGACCCCGACCGCGTTCGGATGGGTGAGAGTGGACATCCGC 70150
QY 1343 TTTCATGATCCACCGGATGACGCGGAATGCGATCCTCAACACGTCGCGGTGACCTCA 1402
DB 70149 TGGCGATGATCTGTTTGA-----CCTGTGACGCTGCGGTGAACCTGG 70105
QY 1403 CCGGTACCCCGCGCTGACGGTGGCCACGGG---TGCGGGGAGAGGCGCTGCCCGTTG 1459
DB 70104 CCGGCACCTGCGGCTATGTTGTCGCTCGGGCTCTCCCGACACGCGGTTCGCGTTG 70045
QY 1460 GCTTCAAGTATAGCGCCCACTTCGAGGAGTCGACGCTCTACCGCACCGCGCGCTGA 1519
DB 70044 GCCTACAGATCATGGCGCGGCAATTGGCGGACGACCGGCTCTACCGGTGCGGGCGGCTT 69985
QY 1520 TCGAGCGCGCGGC 1533
DB 69984 ATGAGCGCGCGGC 69971

RESULT 11

AAT62387
ID AAT62387 standard; DNA; 4730 BP.

XX AAT62387;

DT 24-JUN-1997 (first entry)

XX R. rhodochrous amidase/nitrite hydratase gene containing fragment.

DE Rhodococcus rhodochrous; amidase; probe; amplification; primer; PCR;
KW polymerase chain reaction; phage; open reading frame; nitrite hydratase;
KW enzyme; genetic engineering; ds.

XX Rhodococcus rhodochrous.

XX Key Location/Qualifiers
FH 330..334
FT RBS /tag= a
FT /tag= b
FT CDS /product= "amidase"
FT /note= "this fragment is missing 45 bases between
FT nucleotides 1413-1457 (nucleotides 1069-1113 of the
FT amidase gene (AAT62385))"
FT 1928..1932
FT /tag= c
FT CDS 1939..2562
FT /tag= d
FT /product= "nitrite hydratase alpha subunit"
FT 2589..3227
FT /tag= e
FT /product= "nitrite hydratase beta subunit"
FT terminator 3243..3284
FT /tag= f
FT CDS 3326..4525
FT /tag= g
FT /product= "unknown protein"
FT /note= "this CDS is hypothetical"

PN JP09009973-A.

XX

PD 14-JAN-1997.
XX 27-JUN-1995; 95JP-00184934.
XX 27-JUN-1995; 95JP-00184934.
PR (CHCC) CHISSO CORP.
PA WPI; 1997-126430/12.
XX P-PSDB; AAW14159, AAW14160, AAW14161, AAW14162.
DR Rhodococcus rhodochrous nitrite hydratase and amidase genes - useful for
the industrial preparation of useful cpds.
XX Example; Page 9-15; 16pp; Japanese.
XX This is the nucleotide sequence of a fragment isolated from Rhodococcus
rhodochrous strain IP015564. The sequence was isolated by screening a
genomic R. rhodochrous DNA library in EMBL3 vector using a probe
amplified by the primers T62388-9. The primers amplified a 0.7 kb
fragment of the amidase gene. The library screen isolated a phage
containing a 15 kb insert. When analysed, the insert was seen to contain
an approx. 6 kb SacI-EcoRI fragment containing the open reading frames
for the amidase gene (T62385), the nitrite hydratase alpha and beta
subunits (T62386) and another putative open reading frame encoding a 399
amino acid protein (W14162) of unknown function. The enzymes, prepared by
genetic engineering methods, can be used in the industrial production of
useful compounds
XX SQ Sequence 4730 BP; 1001 A; 1517 C; 1367 G; 845 T; 0 U; 0 Other;

Query Match 10.0%; Score 157.4; DB 2; Length 4730;
Best Local Similarity 53.3%; Pred. No. 2.8e-18;
Matches 379; Conservative 0; Mismatches 326; Indels 6; Gaps 2;
QY 217 CCGACCGCGCAGAACCGGTTCAACGATTCATCCGTTTCGCGGGTGGAGGGCGCC 276
DB 534 CCAACACGCGAGCGAAATCCTTTGAGCGCTTGTGTATGTGACCAACAAGCATCCCGCGCAGC 593
QY 277 ACGAGGGGGCCACTGAGCGACCTGACCGCGCATCAAGGACTGCTATCGGCATTCGCGGT 336
DB 594 TCGACGCGCTCTGACCGCGCGCGCATCAGGACAACTGACCGGTGGCGGA 653
QY 337 ATGCCACACGACGAGCGGTCGCGATGCTCCCGATGTATGCCACCGAGGATGCGGTG 396
DB 654 GTTCCGATGATGAACGGGCTCTCGGACAGTAGAGGGGTTCACTCCGCTCTCGCGACGCACT 713
QY 397 GTGGTGGAGCGGCTGCTCGCGGCGAGCGCCACCATCGTGGCAAGACGACCTCGAGGAC 456
DB 714 GTGATCACTCGACTACTGGCGCGCGGTGCAACCGTTCGGGCGAAGCTGTGTGAGGAC 773
QY 457 ATGGCGAT---GGGTATCGGTGAAGCGAGCGTCTACGGTCTCGGTGTAACCCGACAC 513
DB 774 CTGTGTTTCTCCGTTTCGAGCTTCACACCGCGCAAGCGGACCGGTTCGCAATCCATGGGAC 833
QY 514 CCGCGCCACGCGACGCGGTGATCTTCAGCGGCTCCGGCGCTCCGTCGCTGCGCGCATG 573
DB 834 CCACAGCGTGAAGCAGGTGATCATCCGTTGGCAGTCCGCTCTGTCGCAACAGGTGAC 893
QY 574 GTCGACTTCGCGCTGGCGTTCGATGAGGCGAGGACGATCCGCGATCCGCGCGCATGTGTC 633
DB 894 GTCGATTTTCCCATCGCGCGGATCAGGGTGGATCGATCCGATCCCGCGCATTCGTCG 953
QY 634 GGACTGTTCGCGATGAAGGCGACCCAGCGCTGGTCCGCTTTACGCGCTGACATACATG 693
DB 954 GCGCTGCTCGGACCAAGCGCGATTCGGGCTCGTCCCGTATACCGGTGCAATTCCCATC 1013
QY 694 GACCACACCTTGGACCACTCGGGCCCATCACCAGGGGGGTTCGAGCTCAACGCGCGGTC 753
DB 1014 GAGCGAACATCGACCATCTCGGCCGATCACAGCAGCTCCACGATGCGCGCATGATG 1073
QY 754 CTCGAGGTGTGGCGGGGCGGACTGCGCGGACCCCTCAGTGGGTGCGCTAACCTTCCCGAG 813

Db 1074 CTCCTGGTCTATCGCGGTCTCGGACGCTAGACACCGCCAGACGAGCT---CGAA 1130
QY 814 CCGGAGAACTACGGTCTCCCGCTCGCGAGGAGATATCCGGTCTGAGATTCGGTCTGTC 873
Db 1131 GCAGGTGACTATCTGTCCACCTCGACTCCGATGTGGATGGTCTCGAATCGGGATCGTT 1190
QY 874 GAGGAGTCTACTGAGCGGAACGCTGCGAGCGCGGACGCTATCGCGCGCTTC 924
Db 1191 CGAGAGGTTTCGGGACACGGGTCTCACAGCCCGAGGTGCGAGCGCAGTC 1241

RESULT 12

ACA26208
ID ACA26208 standard; DNA; 1488 BP.

AC ACA26208;

XX 19-JUN-2003 (first entry)

DE Prokaryotic essential gene #7865.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.

XX Burkholderia mallei.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US0009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00949893.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX P-PSDB; ABU22336.

XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 14078; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX compound's activity; (11) a culture comprising strains in which the gene
XX product is overexpressed or underexpressed; (12) determining the extent
XX to which each of the strains is present in a culture or collection of
XX strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. Note: The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 1488 BP; 234 A; 520 C; 525 G; 209 T; 0 U; 0 Other;

Query Match 9.6%; Score 152.2; DB 7; Length 1488;

Best Local Similarity 48.6%; Pred. No. 2.3e-17;

Matches 631; Conservative 0; Mismatches 623; Indels 45; Gaps 6;

QY 273 CGCCACGAGGGGCGGACCTGAGCGACCTGACCGCGCATCAAGCATGCTCGCCATCGC 332

Db 193 CGGCGAGCGGGGCGGCTCGCGGCTGCGCATCGTGCACAAGACGTGTGTGTCAGCG 242

QY 333 CGGTATGCCACACGAAACGGGTCCCGATGCTCCGACTGTGATCGCCACCGAGGATGC 392

Db 243 CGGTGGGCTCCACCGCGGCTCGAAGATGCTCGGAATACGCGAGCGGCTTCGACGC 302

QY 393 CGTGTGTGAGCGGCTGTCGCGGAGCGGCCACCATCGTGGGAGAGACGACCTCGA 452

Db 303 GACCGTGTGCGCGGCTGTCGCGGCGCGCATGCGTCAAGCATGGA 362

QY 453 GGACATGGCGATGGGT---ATCGGTGAAGGAGGAGGCTTACGGTCTCTGCGTGAACCCGAA 509

Db 363 CGATTTCGGATGGGTTCGTGCAAGAGACTCCGCTTCGGCCCGGTGAAGAACCCGTG 422

QY 510 CAACCCGCCCGACGCGAGCGGTGATCTTCAGCGGCTTCGGCGCTGCGTTCGTCGCGG 569

Db 423 GGACACGAGCGCGGTGCGCGGCGGAGCTCCGCGCGGCTGCGCGCGG 482

QY 570 CATGTGACATTCGCGCTGCGGCTGCGATGAGGAGGAGGATCCGGAATCCGCGCGCATG 629

Db 483 CTTGCGCGCGCGGAGCGGACGACGACACGCGGCGGTTCGATCGCCAGCGCGCTGCTT 542

QY 630 GTGCGGACTGTGCGCATGAAAGCGAACCCAGCGCTGTGTCGCTTACGGCTTGACATA 689

Db 543 CGCGCGGTGACCGGCTCAAGCGGACCTACGCGCGGTGTCGCGCTACGGGATGATCGC 602

QY 690 CATGACACACCTTGGACACATCGGGCCCATCACGAGGGGGTTCGAGCTCAACGCCCG 749

Db 603 GTTCGCGTGTGCTGCTGACCGAGCGCGGCGGATGCGCGGAGCGCGGCGGCTTGGCG 662

QY 750 GGTCTCTGAGGTGTTGGCGCGGCGGACTGCGCGGACCTCAGTGGGT----- 797

Db 663 CTTGCTGAACGCGATGGCGGCTTCGACGAGCGGCTTCGACGAGCGCGCGCA 722

QY 798 -----GGTAACTTCGGAGCGGAGAACTACGGCTCCGCTCGCGGAGGG 845

Db 723 CGAAGACTACACGCCCGGCTTCGGCAAGGCTTGGCGCGCGGCGGAGCGCGGCGCAAGCC 782

QY 846 AGTATTCGCTGTAGATTTCGGGTCTGCGAGGAGTCACTGGAGCGCGAAACGGTTCGACGCC 905

Db 783 GTCGCGGCGCTGCGATCGGCTTCGCGCGGATATTCGCGCGGCGGCTTCGCGCA--- 839

QY 906 GGAAGTGTGCGCGGCTTCAACGAGGAGTGTGCGCGGCTTCGAGAGCGCGGCTGCGACCAT 965

Db 840 CGACGTGCGCGCGGCGGCTGCGAGCGCGCTTGAAGACCTTACGAGGCGGCTTCGCGCGGCG 899

QY 966 CGAGCGGCTTCGCGTGGCTGTCGAGCGCGGCTTCGAGAGCGCGGCTTCGAGAGCGCGGCTGATGGC 1025

Db 900 CGTTCGCGGTGCTGCTGCGGAGACGAGGCTGTGCTGATTCGCTTACTATGTGATCGCGC 959

QY 1026 TTTCACGCGCGGCTA---TGGCGGACTCCCGCGGTGTGGCTACTTCCACAAAGGGGG 1082

Db 960 CGCGGAGGCGTGTGCGAAGCTTGTGCGGCTTCACGCGCGTGTGCGCTACGGCCAC-----CG 1013

QY 690 CATGGACACACCTTGGACACATCGGGCCCATCCAGGGGGTGCAGCTCAAGCCCG 749
Db 609 CTACGCTTCAGGCTCGACAGGGCGCCGCTGGCGGCACACCGCGAGCTGCGCGCT 668
QY 750 GGTCTCTCGAGGTGTTGGCGGGCCGAGCTGGCGGACCTCGAGTGGGTGCGTAACCTCC 809
Db 669 GATGCTGGGGGTGATGGCGGATTCGATCCGAAGGACTCGACAGCGT-----CGACA 722
QY 810 GGAGCCGGAGAACTAGCGCTCGCGCTCGCGGAGGAGTATCGGCTCGAGATTGCGGT 869
Db 723 GCGGTGGAGCACTACTGCGCCGCTCGAGAGCCGCTGAGCGGCTCGCATCGGCT 782
QY 870 CGTCGAGGATCTACTGAGCGGAACGCTGCGACCGCGGACGTGATCGCGCGCTTCAACCA 929
Db 783 GCGCGGGAATACTTGGCGCGGCTCGACAGCCGCATCGCGACGGGTGCTGCGCGT 842
QY 930 GGAGCTGGCGGCTCGAGAGCGCGGTGCGGACCATCGAGCGGTCTCGGTGCGGTG 989
Db 843 GG---TCGAGAGCTGAAGAGCGTTCGCGCGCCAGGTGAAGGACATTTCCCTCCGAACAT 899
QY 990 GAGCGGCGCTGCGCTATCCAGAGCGCGGTGATGGGTTTCAACGCGCGGCTATGCGGA 1049
Db 900 GCAGCAGCCATCCGCGCTACTAGTATCGCGCGCGGAGGCTCTCCACCTGTC 959
QY 1050 CTCGCGGTGGGTCTTCCACAGGGCGCGGTGAGCGTACGACCGCGCTCAGCAC 1109
Db 960 GCGCTTGCAGCGGCTGCGTATGGCTATCGTTGCGAGCGCCCGCAGAACCTGGAAGACCT 1019
QY 1110 GCGCGCCGAGTGCACCAACCCACAGGATCTGGCGATCTGTCGCGGCTGATGCTGT 1169
Db 1020 GTACAGGCTCGCGCGGAGGCTTCGCGAGCGAGTGAAGAACCGCATATGTCGG 1079
QY 1170 GATCGCGGACACTGCGGACGAATACCTTCGCGATCCACTACGGAAGCGCAGAACCT 1229
Db 1080 CACCTACGCACTCTCG---GCCGGCTACTACGATGCTATTACTCGAGGCTCAGAAGAT 1136
QY 1230 GCGGCTGGAGCTCGGAAGCAGATCGACGCGCTCTCCAGGACCGGCTGCACTGTCAC 1289
Db 1137 TCGCCGCTGATCAAGAACGACTTCGTACGCGCTTTCGCGAGTGGAGTATGCTCGG 1196
QY 1290 CCGGACACGCTTACCGTTGCCAACGAGCTGTTGAGCGGTGCGCAAGACACCATGTCAT 1349
Db 1197 CCGGACACGCGCAACCGCGCTGGAAGATCGCGGAGAGAACGACGACCCGTTTCCCA 1256
QY 1350 GATCCACGATCAGCGGCAATCGATCTCAACAGTGCCTGAGCTACCGGTCA 1409
Db 1257 GTACCTG-----GAAGACATCTACACCATCACCGCAACCTCGCGGCT 1301
QY 1410 CCGCGGCTGACGCTGCCACCGGTCGCGGCGAGAGGCGCTCGCGGCTTGGCTCCAAGT 1469
Db 1302 GCGGGGCTGTCCATGCGCGCGGCTTGTGTCG---ACGGCTGCGGTGCTCAGTT 1358
QY 1470 GATAGGCGCACTTCAGGA 1490
Db 1359 GCTCGCGCCCTACTTCCAGGA 1379

RESULT 15

ABQ81844/c

ID ABQ81844 standard; DNA; 349980 BP.

XX

XX

AC ABQ81844;

DT 19-NOV-2002 (first entry)

XX

XX

DE Bifidobacterium longum NCC2705 related nucleotide sequence SEQ ID:1100.

XX

KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;

KW antidiarrhetic; antibacterial; inhibitor of Salmonella; detection;

KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;

KW rotavirus; food composition; pharmaceutical composition; gene; ds.

XX

OS Bifidobacterium longum.

Synthetic.

OS EP1227152-A1.

XX 31-JUL-2002.

XX 30-JAN-2001; 2001EP-00102050.

XX 30-JAN-2001; 2001EP-00102050.

XX (NEST) SOC PROD NESTLE SA.

XX WPI; 2002-668397/72.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

OS The present invention describes a polynucleotide (I) comprising a sequence of a Bifidobacterium genome selected from the nucleotide sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at least 90% identity or which hybridizes with the sequences given in ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a fusion protein, comprising a sequence selected from 1097 sequences given in ABQ81842 and ABQ81843 ligated in frame to a polynucleotide encoding a heterologous polypeptide. (I) has antidiarrhetic and antibacterial activities, and can be used as an inhibitor of Salmonella. (I) (which is a probe) is useful for the detection and/or identification of Bifidobacterium longum in a biological sample. A carrier containing the lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM I-2618) can be used for preventing and/or treating diarrhoea brought about by pathogenic bacteria and/or rotavirus. The carrier is a food composition selected from milk, yogurt, curd, cheese, fermented milks, milk based fermented products, ice-creams, fermented cereal based products, milk based powders, infant formulae, pet food or a pharmaceutical composition selected from tablets, liquid bacterial suspensions, dried oral supplement, wet oral supplement, dry tube feeding or wet tube feeding. (I) is useful in DNA arrays or chips to carry out analysis of the expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent Bifidobacterium related nucleotide sequences given in the sequence listing from the present invention but not mentioned further within the specification. N.B. The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied by the European Patent Office

Sequence 349980 BP; 69975 A; 105045 C; 104394 G; 70566 T; 0 U; 0 Other;

Query Match 9.3%; Score 146.4; DB 6; Length 349980;
Best Local Similarity 47.0%; Pred. No. 1.8e-16;
Matches 681; Conservative 0; Mismatches 731; Indels 36; Gaps 6;

QY 100 GCCGCCAGCTCGTTCGCGGTGTCGGCGAGATGGTACTGCTTTCGACCTGATCGAGCA 159
Db 211290 GCCGAGCGCGCTGCGGTCAAGAGCGGAGCTCACCTCTCGCGAGCTCGTGAAGCC 211231
QY 160 CTACCGCAACCGCGCGAGCCCGGACCGGTACAGGACCGCGGACATCGCGCGGACCG 219
Db 211230 CACCTCAAGGTGATCGAGGCTGCGGAGCTCCATCAAGGCTTCTCTGAAAGTTTCGGC 211171
QY 220 ACCGCGGACGAGACCGGTTCAACGATTTCATCGGTTCTCGCGGTGAGGGGCGCCACG 279
Db 211170 GATGTGCTCTCGAGCGAGCGGCGCTTCGAGCCCAAGTCCGCGGAGACAGGCGCC 211111
QY 280 GAGGGGCGATGAGCGGACCTGACCGCGCGATCAAGGACTGATCGCCATCGCGGTATG 339
Db 211110 CTGCGCTGAGCTGGTGGTGTGCGGATCGCCATCAAGGACATGATCGTCCAGGGGATC 211051
QY 340 CCCACCGAGAGCGGTCCCGGATCTCCGACTGTGATCGCACCGAGGATGCGGTGTG 399
Db 211050 GAAACCGCGCGCTTCAAGATCTTGAAGGTGGGTGCGCGTACGACGCGCCGTC 210991

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 04:30:32 ; Search time 4525.69 Seconds

(Without alignments)
10412.231 Million cell updates/sec

Title: US-10-658-691-4

Perfect score: 1578

Sequence: 1 gtgcgccaatgccatt.....gcgcacccgtgctgcggtag 1578

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estnu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_estc.*

9: gb_est1.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_lman.*

23: em_gss_lmus.*

24: em_gss_pro.*

25: em_gss_rod.*

26: em_gss_phg.*

27: em_gss_vrl.*

28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	147	9.3	540	28	BZ893076 HL11_0118
2	145.6	9.2	600	28	BZ895409 NaRPto_01
3	131.6	8.3	583	13	BQ245721 TaE15021A
4	122.6	7.8	688	14	CB661518 OSJNEd04H

5	118.6	7.5	517	10	BZ426574
6	117	7.4	636	10	BZ426316
7	114.2	7.2	659	28	BZ891923
8	113.4	7.2	857	10	BF262005
9	112.4	7.1	535	14	CD874694
10	111.4	7.1	640	28	AZ935507
11	107.6	6.8	604	28	AZ935188
12	106.6	6.8	2106	28	AQ012146
13	106.2	6.7	521	28	AZ934731
14	106	6.7	672	28	AZ934200
15	104.6	6.6	517	12	EG300285
16	104.2	6.6	506	14	CA685161
17	104	6.6	595	28	AZ934675
18	101.2	6.4	722	14	CB632529
19	100.2	6.3	753	14	CA233999
20	100	6.3	696	13	BZ896649
21	100	6.3	829	13	BZ118499
22	97.8	6.2	668	14	CA500482
23	96	6.1	588	14	CD208439
24	94.6	6.0	1313	28	BZ548424
25	94.2	6.0	816	14	CB659357
26	91.6	5.8	451	14	CD873667
27	91.2	5.8	515	14	CF646949
28	90.8	5.8	440	14	CB641112
29	90.6	5.7	937	14	CA289259
30	90	5.7	677	14	CA290026
31	89.6	5.7	558	10	AW564214
32	89.6	5.7	571	28	AZ934852
33	89	5.6	749	13	BQ840610
34	88.4	5.6	599	14	CA655576
35	88.2	5.6	412	28	BZ345145
36	85.8	5.4	632	14	CB851355
37	83.2	5.3	719	9	AJ612670
38	83	5.3	445	10	BE500531
39	83	5.3	742	14	CB630965
40	82.6	5.2	662	12	BM426394
41	82.2	5.2	619	9	AL504794
42	81.8	5.2	701	14	CB853233
43	81.8	5.2	773	14	CD433011
44	81	5.1	711	14	CA289750
45	78.8	5.0	814	28	BZ578429

ALIGNMENTS

RESULT 1
BZ893076
LOCUS HL11_0118 HL pUC18 Library Halorubrum lacusprofundi genomic 5',
DEFINITION HL11_0118 HL pUC18 Library Halorubrum lacusprofundi genomic 5',
GENOMIC SURVEY SEQUENCE.
ACCESSION BZ893076
VERSION BZ893076.1 GI:33343666
KEYWORDS GSS.
SOURCE Halorubrum lacusprofundi
ORGANISM Halorubrum lacusprofundi
Archaea: Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halorubrum.
REFERENCE 1 (bases 1 to 540)
AUTHORS Goo,Y., Roach,J., Glusman,G., Baliga,N.S., Deutsch,K., Fan,M.,
DasSarma,S., Ng,W.V. and Hood,L.
TITLE Low-pass Sequencing for Microbial Comparative Genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Goo Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoo@systemsbiology.org
Seq primer: M13 Forward
Class: shotgun.
Location/Qualifiers
1..540

FEATURES
source
1..540

```
/organism="Halorubrum lacusprofundi"
/mol_type="genomic DNA"
/strain="ATCC 49239"
/db_xref="taxon:12247"
/clone_lib="HI pUC18 Library"
/notes="Vector: pUC18; Site 1: SmaI; A shotgun library was
constructed from Halorubrum lacusprofundi genomic DNA
using pUC18/SmaI/BAP plasmid"

ORIGIN
Query Match      9.3%; Score 147; DB 28; Length 540;
Best Local Similarity 58.9%; Pred. No. 6.9e-12;
Matches 271; Conservative 0; Mismatches 186; Indels 3; Gaps 1;

QY 330 CGCGGTATGCCACCAACGCGTCCCGGATGCTCCCGACTGTGATCGCCACGAGGA 389
Db 45 CGCGGGATCGGACACCTGGGCTCCGAGATGCTCGCGACTAGTCCCGCTATC 105
QY 390 TCGCGTGGTGGAGCGCTCTCCGCGAGCGCCACCATCTGTCGGCAAGCAACT 449
Db 106 CGCGACTGTCTGACCGGCTGACCGACGCGGCGCGACGCTGCGGCAAGCAAT 165
QY 450 CGAGGACATGGCGATGGG---TATCGTGAAGCAGCTCTACGGTCTTCGCTCAACCC 506
Db 166 GGACGAGTCTTTATGGGAACGACACCGAGACTCCGGTTCGGCCCAAGAACCC 225
QY 507 GAACAACCCCGCCACGCGAGGTGATCTTCCAGCGCTCCGCGGTTCGCTGTC 566
Db 226 TCCGACCCGAGCGGCTCCCGCGGCTCCTCGGCGGCTCGGCGGCGCTCGCGC 285
QY 567 CGCATGTGCTGACTTCGCTCGGCTGATGATGAGGAGCAGCAGCAGTCCGGATCCCGCGC 526
Db 286 GGGCGAGGCGAGCTGGCTCGCTCCGACACGCGCGGTGCGTGGTTCGCGCGC 345
QY 627 ATGTCGCGACTGTCGCGATGAAGCGACCCACCGCTGTCGCTTTACGGCTGAC 686
Db 346 GTTCTCGGCGTCTCGGATCAAGCCACCTACGGGCTCGTCTCGCGTACGGCTCGT 405
QY 687 ATACATGACCCACACTTGGACCATCGGCGCCATCACCAGGGGTGAGCTCAAGC 746
Db 406 CGCGTACGAACTCCCTCGACAGATCGGACCGATCGCGGACCGTTCGAGGCGGC 465
QY 747 CGGGTCTCGAGTGTTCGCGGCGCGACTGGCGCGAC 786
Db 466 CGNCGCCCTGACGTATCGTGTGAGCGGACCGCAGCAGAC 505

RESULT 2
BZ895409
LOCUS
DEFINITION
NaP10.0171 Na pUC18 Library Natrialba asiatica genomic 5', genomic
survey sequence.
ACCESSION
BZ895409
VERSION
BZ895409.1 GI:33345885
KEYWORDS
GSS
SOURCE
Natrialba asiatica
ORGANISM
Natrialba asiatica
Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
Halobacteriaceae; Natrialba.
REFERENCE
1. (bases 1 to 600)
Goo Y., Roach J., Glusman G., Baliga N.S., Deutsch K., Pan M.,
DasSarma S., Ng W.V. and Hood L.
Low-pass Sequencing for Microbial Comparative Genomics
Unpublished (2003)
Contact: Goo Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoo@systemsbiology.org
Seq primer: M13 Forward
Class: shotgun.
Location/Qualifiers

FEATURES
Location/Qualifiers
```

```
1. 600
/organism="Natrialba asiatica"
/mol_type="genomic DNA"
/strain="ATCC 700177"
/db_xref="taxon:64602"
/clone_lib="Na pUC18 Library"
/notes="Vector: pUC18; Site 1: SmaI; A shotgun library was
constructed from Natrialba asiatica genomic DNA using
pUC18/SmaI/BAP plasmid"

ORIGIN
Query Match      9.2%; Score 145.6; DB 28; Length 600;
Best Local Similarity 55.7%; Pred. No. 1.1e-11;
Matches 324; Conservative 0; Mismatches 249; Indels 9; Gaps 2;

QY 407 GGCTGCTCGCGGAGCGCCACCATCTGTCGGCAAGACGACCTCGAGGACATGGCGATGG 466
Db 1 GGCTCAAAGACGCGCGCGACCATCTGTCGGCTTTTACCAACATGCTTTTTCGGCATGG 60
QY 467 GT---ATCGGTGAAGGAGCGTCTACGGTCTCGCTGTGAACCCGAAACCAACCCGCCACG 523
Db 61 GTACGACACCGAGACTCTCGCTTCGGTCCGACCGACACCCCGCGCGCGGCCACG 120
QY 524 GCAGGCTGATCTTCAGCGGCTCCGCGCTCGCTGCTCGCTCGCGATGGTTCGACTTCG 583
Db 121 TTCCCGCGCTCTCTCGGCGGCTCAAGCCGCGCTCGCGCGCGGCGAAGCGGAACTCG 180
QY 584 CCCTGGCGCTCGATGAGCGAGGAGCATCCGATCCCGCGCGCATGCTGCGGACTGCTCG 643
Db 181 CGCTCGCTCGATACGGCGGCTCGATCGCTGTCCGCGCGCTGCTGCGGCTCGTTCG 240
QY 644 GCATGAAGGCGACCCACGCGCTCGTTCGCTTTACGGCTTGAATATGAGACACACCT 703
Db 241 GGATCAAGCGGACTTACGGCTCTGTCGCGCTACGGCTCGTTCGCTACGCAAGCGCC 300
QY 704 TGAACCATCTCGGCGCCATCACCAGGGGTCGAGCTCAACCGCCGCTCTCGAGTGT 763
Db 301 TAGAGAGATCGGCGCTTCGCGAGACCGTCAAGACGCGCGCGCTCTCGACTCA 360
QY 764 TGGCGGCGCGACTGCGCGACCCCTCAGTGGTGTGTAACCTTCGCGAGCGCGAGAACT 823
Db 361 TCGCAGACCGCGACGAAACGAGCGCGACTCGAAGACCGAACTCGAGGGGCGAAGCT 420
QY 824 ACGGCTCGCGCTCGCGGAGGATATCGGCTCGATTCGCGCTGCTGAGGAGTTCAC 883
Db 421 ACGCCGACCGCGCAACCGCGCGCTTACGATCGCGCTTACGATCGCGCTTCCACCGCACTGG 480
QY 884 TGGAGCGCAACGGTTCGCGACCGCGCGCTTCAACCGAGGAGCTCGCGCGCGC 943
Db 481 TCGAG-----GGCGCGACGACGCGCTCTCGACACCTTCTGGGACGCCATCGCTGAAC 534
QY 944 TCGAGAGCGCGCTGCGACCATCGAGGGCTCTCGGTGCGGT 985
Db 535 TCGAGGCCCGCGCGCGAGTATCAGAGTCTCGCTACCGT 576

RESULT 3
BQ245721
LOCUS
DEFINITION
TaE15021A02R TaE15 Triticum aestivum cDNA clone TaE15021A02R, mRNA
sequence.
ACCESSION
BQ245721
VERSION
BQ245721.1 GI:20441597
KEYWORDS
EST.
SOURCE
Triticum aestivum (bread wheat)
ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE
1. (bases 1 to 583)
Cloutier S.
Wheat functional genomics - Glenlea developing seeds cDNA libraries
Unpublished (2002)
JOURNAL
```

COMMENT

Contact: Dr. Sylvie Cloutier
Cereal Research Centre, Agriculture and Agri-food Canada
195 Dufour Rd, Winnipeg, MB, Canada R3T 2M9
Tel: (204) 983-2340
Fax: (204) 983-4604
Email: scloutier@agr.gc.ca

was cloned directionally, not all sequences generated with reverse primer were from the 5' end (same with forward primer and 3' end).
Average insert size is >1.4 kb
Plate: 021 row: A column: 02
Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers
1..583
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Glenlea"
/db_xref="taxon:4565"
/clone="TaB15021A02R"
/tissue_type="developing seeds"
/dev_stage="15 days after anthesis"
/lab_host="E. coli DH10B"
/clone_lib="TaB15"
/note="Vector: pCMV-SPORT6.0 (Invitrogen Technologies); Site 1: NotI; Site 2: MluI; mRNA obtained from wheat seeds of cultivar Glenlea 15 days post-anthesis"

ORIGIN

Query Match 8.3%; Score 131.6; DB 13; Length 583;
Best Local Similarity 58.1%; Pred. No. 1.2e-09;
Matches 274; Conservative 0; Mismatches 189; Indels 9; Gaps 2;
QY 262 CCGGTGGAGGGCGCACGAGGGGCGACCTGAGCGACCTGACCGCGCGCATCAAGACTGC 321
DB 103 CGCGAGCAGCCCGCCCTGACGCGCTCACCTTCGCCATCAAGGACATCTCGAGCTGCC 162
QY 322 ATCGCCATCGCGGTATGCCACCAAGAGCGGTCCCGGATGCTCCGACGTGATGCC 381
DB 163 GCGCGGTACCGGTTTGGACACCGGACTGGCGCGGACGCGCGCGCGCC 222
QY 382 ACCGAGATGCGGTGGTGGAGCGGTGCTCGCGGAGGCGCCACCATCGTGGCAAG 441
DB 223 ACGGCGCGCGCGTGTGCGCGCGTG-----GCCGCGGCGCCACGCGGTGGGCAAG 276
QY 442 ACGAACCTCGAGACATGCGATGGTAT---CGGTGAAGCAGCGTCTAGCGTCTGCG 498
DB 277 ACGGTGATGGAGATGCGCTTACAGCATCAACGGCGAGAACGCGCATACGCGACCCCC 336
QY 499 CTGAACCCGAAACACCCCGCCCGCGCGCTCCCGCGGCTCCCGCGCTGCC 558
DB 337 GCCAACCCCTGCGCGCGCGCGCTCCCGCGGCTCTCCAGCGGTCCCGCGTCCG 396
QY 559 GTGCTGCGCGCATGCTGCACTTCGCTCCGCTGCGCTGATGAGGAGGAGCATCGGATC 618
DB 397 GTGCGCGCGAGCTCGCGACTTCGCGCTCGGCACTGGCGGCGAGCGTCAAGGTG 456
QY 619 CCGGCGCATGTGGGAGTGTGCGGATGAAGGAGACCCAGCGCTGCTGCGCTTAC 678
DB 457 CCGCGCGCATCTGCGGCATCTTGGCTCCGCTCCGCGCTCCAGGACTGCTCCACCGAG 516
QY 679 GGCCTGACATATGAGCACACCTTGGACACATCGGGGCCCATCACCAGG 730
DB 517 AACGTGTCGCCATGGCGGAGATGTCGACACTGTGCGGTGTTGCTAGGG 568

RESULT 4

CB661518

LOCUS

DEFINITION CB661518 688 bp mRNA linear EST 09-APR-2003
clone OSJNEd04H23.f OSJNEd Oryza sativa (japonica cultivar-group) cDNA

ACCESSION

VERSION

KEYWORDS

SOURCE Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

AUTHORS

1 (bases 1 to 688)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,B.,
Kidnra,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)

JOURNAL

COMMENT

Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088 USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gtg
BACKWARD: gga aac agc tat gac cat g
Plate: 04 row: H column: 23
Seq primer: gta aaa cga cgg cca gtg.

FEATURES

Location/Qualifiers
1..688

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/clone="OSJNEd04H23"

/tissue_type="leaf"

/dev_stage="3 week"

/lab_host="DH10B"

/clone_lib="OSJNEd"

/notes="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 24 hrs after inoculation with Rice Blast (C9240-1)"

ORIGIN

Query Match 7.8%; Score 122.6; DB 14; Length 688;
Best Local Similarity 58.7%; Pred. No. 2.7e-08;
Matches 253; Conservative 0; Mismatches 169; Indels 9; Gaps 2;
QY 286 CCACTGAGCGACCTGACCGCGCGGATGCTCCGACTGTGATCGCCACCGAGGATCGCGTGTG 345
DB 96 CCCCTGACGGCTCACCTTCTCATCAGGACATTTTCGACATCGCGGGCGGTGACC 155
QY 346 ACG-----AACGGTCCGGATGCTCCGACTGTGATCGCCACCGAGGATCGCGTGTG 399
DB 156 GGGTTTCGCAACCGGACTGGCGAGGACCCAGCGCGCGCGCGCCACCTCCCGGTC 215
QY 400 GTGAGCGGTGCTCGCGGAGCGGCGCACCATCGTCGGCAAGACGACCTCGAGGACATG 459
DB 216 GTCTCGCGCGCTCGCGCGCGGCGCACCGCTCGGACACCATCATGAGACGAGATG 275
QY 460 GCGATGCGGTATCGGTGAAGGACGCG---TCTACGCTCTGCTGAAACCGAAACACCCC 516
DB 276 GCATACAGCATCTATGGCGAAGACGCGCATACGCGACGCGCGCGCGCGCGCGCC 335
QY 517 GCCCAGGCGGTGATCTTCAGGGCTCGCGGCTCGCGTCTCGCTGCGTGGCGCATGTC 576
DB 336 GGCAGAGTCCCGCGGAGATCTCCAGTGGCTCGCGCTCGCGTTCGCGCGCAACCTGTC 395
QY 577 GACTTCGCCCTCGGCGTTCGATGAGGAGGAGGAGCATCCGATCCCGCGCGCATGTCGCGGA 636
DB 396 GACTTCTCTCTCGGACCGACACCGCGCGGAGGAGGAGTGCCTGCTACTGCGGC 455
QY 637 CTGCTCGGCTAGAGGCGACCCACCGGCTGTCGCTTTACGGCTGACATACATGAC 696
DB 456 ATCTACGGCTTCGCGACCTCCCATGCGCTCGTTTCGCTCAAAACGTCATCCCATGAA 515
QY 697 CACACCTTGA 707
DB 516 CAAATGTGTGA 526

```

RESULT 5
BE426574
LOCUS
DEFINITION
  BE426574.1 517 bp mRNA linear EST 24-JUL-2000
  Triticum aestivum cDNA clone WHE0336_H04_P08, mRNA sequence.
ACCESSION
  BE426574
VERSION
  BE426574.1 GI:9424417
KEYWORDS
  EST
SOURCE
  Triticum aestivum (bread wheat)
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Pooidae; Triticeae; Triticum.
  1 (bases 1 to 517)
REFERENCE
  Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D.,
  Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,
  Seaton, C.L. and Tong, J.C.
  The structure and function of the expressed portion of the wheat
  genomes
  Unpublished (2000)
  Contact: Olin Anderson
  US Department of Agriculture, Agriculture Research Service, Pacific
  West Area, Western Regional Research Center
  800 Buchanan Street, Albany, CA 94710, USA
  Tel: 5105595773
  Fax: 5105595818
  Email: canderan@pw.usda.gov
  Sequence have been trimmed to remove vector sequence and low
  quality sequence with phred score less than 20
  Seq primer: Stratagene SK primer.
  Location/Qualifiers
    1..517
      /organism="Triticum aestivum"
      /mol_type="mRNA"
      /cultivar="Chinese Spring"
      /db_xref="taxon:4565"
      /clone="WHE0336_H04_P08"
      /tissue_type="Etolated shoot"
      /dev_stage="Five day old seedling"
      /lab_host="E. coli SOLR"
      /clone_lib="Wheat unstressed seedling shoot cDNA library"
      /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
      Site 1: EcoRI; Site 2: XhoI; Seeds were
      surface-sterilized, germinated and grown aseptically in
      the dark at room temperature on filter paper with water,
      mystatin and cefotaxime in covered crystallization
      dishes. Shoots were harvested. The tissue, total RNA, and
      the cDNA clones were in vivo excised to give pBluescript
      phagemids in the TJ Close lab (Choi, Close, Fenton) at the
      University of California, Riverside. Plasmid DNA
      preparations and DNA sequencing were performed in the OD
      Anderson lab (all other authors)."
    7.5%; Score 118.6; DB 10; Length 517;
    Best Local Similarity 60.6%; Pred. No. 9.5e-08;
    Matches 231; Conservative 0; Mismatches 144; Indels 6; Gaps 2;

QY 290 TGAAGACCTGACCGCGGATCAGGACTGATGCCATGCCCGTATGCCACCAAGCA 349
DB 52 TGGACGGCGTGTGTGGCGGTGAAGACGAGATGCTTGCCTTACCCACCAAG 111
QY 350 ACCGGTCCCGATGTCCTCCGACTGTGATGCCACGAGGATGCCGTGTG--GTGGAGC 406
DB 112 GCGGACGCGGTGGTGGGAGGCGCGCCGTGCGAGCGGACGCGCGTGGTGGCGC 171
QY 407 GCGTGTCCGCGACGCGCCACCATCGTGGCAAGACCAACTCGGAGACATGCGCATGG 466
DB 172 AGCTGCGCGTGGCGCGCGTCTCGCGGCAAGACCAACATGACGAGTCGCGCGCG 231

```

```

QY 467 GTATCGGTGAAGGACG---GTCTAGGCTCTGCGCTGACCCGACACCGCGCCACG 523
DB 232 GCACGAGCGCATCAACCCACACCGATCGGAGGAACCCGTACAACTCGGCAAG 291
QY 524 GCACGCGGTGATCTTCCAGCGGCTCGCGGCTCGCGCTCGCTCCGCGCATGGTCTG 583
DB 292 TGGCTGGGGGCTCTCTCAGCGGCTCGGCGCGGTGTGTGTGTGTGTGTGTGTGTGT 351
QY 584 CCTGGGCGCTCATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 643
DB 352 CGCTCGCGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 411
QY 644 GCATGAAGGCGACCCACGCGCC 664
DB 412 GCTTCAAGCCACCGCGCGGAC 432

RESULT 6
BE426316
LOCUS
DEFINITION
  BE426316.1 636 bp mRNA linear EST 24-JUL-2000
  Triticum aestivum cDNA clone WHE0330_C03_E06, mRNA sequence.
ACCESSION
  BE426316
VERSION
  BE426316.1 GI:9424159
KEYWORDS
  EST
SOURCE
  Triticum aestivum (bread wheat)
ORGANISM
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Pooidae; Triticeae; Triticum.
  1 (bases 1 to 636)
REFERENCE
  Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D.,
  Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Rausch, C.J.,
  Seaton, C.L. and Tong, J.C.
  The structure and function of the expressed portion of the wheat
  genomes
  Unpublished (2000)
  Contact: Olin Anderson
  US Department of Agriculture, Agriculture Research Service, Pacific
  West Area, Western Regional Research Center
  800 Buchanan Street, Albany, CA 94710, USA
  Tel: 5105595773
  Fax: 5105595818
  Email: canderan@pw.usda.gov
  Sequence have been trimmed to remove vector sequence and low
  quality sequence with phred score less than 20
  Seq primer: Stratagene SK primer.
  Location/Qualifiers
    1..636
      /organism="Triticum aestivum"
      /mol_type="mRNA"
      /cultivar="Chinese Spring"
      /db_xref="taxon:4565"
      /clone="WHE0330_C03_E06"
      /tissue_type="Etolated shoot"
      /dev_stage="Five day old seedling"
      /lab_host="E. coli SOLR"
      /clone_lib="Wheat unstressed seedling shoot cDNA library"
      /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
      Site 1: EcoRI; Site 2: XhoI; Seeds were
      surface-sterilized, germinated and grown aseptically in
      the dark at room temperature on filter paper with water,
      mystatin and cefotaxime in covered crystallization
      dishes. Shoots were harvested. The tissue, total RNA, and
      poly(A) RNA were prepared, a cDNA library was made, and
      the cDNA clones were in vivo excised to give pBluescript
      phagemids in the TJ Close lab (Choi, Close, Fenton) at the
      University of California, Riverside. Plasmid DNA
      preparations and DNA sequencing were performed in the OD
      Anderson lab (all other authors)."
    7.4%; Score 117; DB 10; Length 636;

ORIGIN
  Query Match

```

Best Local Similarity 60.4%; Pred. No. 1.7e-07;
Matches 230; Conservative 0; Mismatches 145; Indels 6; Gaps 2;

```

QY 290 TGAGCGACCTGACCGCGCGATCAAGGACTGATCGCCATCGCGGTATGCCACCAAGA 349
  |||
Db 39 TGAACGGCTGTGTGGGGTGAAGACGAGATGACTGCTGCCTACCCACCAAG 98
  |||
QY 350 ACGGGTCCGGATGCTCCGATGCTGTGATGCGCACGAGATGCCGTGTG---GTGAGC 406
  |||
Db 99 GCGGACGCGGTGGCTGGGGAAGGCGCGCCCTGCGAGCGGACGCGCGGTGGGCGC 158
  |||
QY 407 GGCTGCTCGCGGACGCGCCACATCTGTCGGCAAGACGAACCTCGAGGACATGGCGATGG 466
  |||
Db 159 ACTGCGGCGATGCGGGCGCGCTCTCGCGGCAAGCAACATGACGAGCTCGGCGCG 218
  |||
QY 467 GTATCGGTGAAGCGAGCGT---CTACGGTCTGCGCTGAACCGCAACACCGCCACG 523
  |||
Db 219 GCACGAGCGGATCAACCGGACCAACGATCGGCGGAGGAACCGTACCAACGTCGCGAAG 278
  |||
QY 524 GCACGCGGTGATCTCCAGCGGCTCGGCGCTCGCGTCTGCGTCCGCGCATGGTCTGACTTCG 583
  |||
Db 279 TGGCGGGGGCTCTCCAGCGGCTCGCGCGCTCGCGTCTGCGTCTGCGCTCTGCGCGCTCG 338
  |||
QY 584 CCTGCGGCTGATGAGGAGGAGCAGCATCCGGATCCCGCGCGCATGGTCCGAGCTGGTCG 643
  |||
Db 339 CGTCTGCGCTGACGAGGAGGCTCTGTCGGATGCGCGCAGCTCTGTGCGGCGGTGTCG 398
  |||
QY 644 GCATGAGCGGACCCACGCGCC 664
  |||
Db 399 GCTTCAAGCCACCGCCGAC 419
  |||

```

RESULT 7
BZ891923/c 659 bp DNA linear GSS 30-JUL-2003
LOCUS Emio_0193.xl_088.ab1 Hm pUC18 Library Haloarcula marismortui
DEFINITION Genomic 5', genomic survey sequence.

ACCESSION BZ891923
VERSION BZ891923.1 GI:33342556
KEYWORDS GSS.
SOURCE Haloarcula marismortui
ORGANISM Haloarcula marismortui
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Haloarcula.
GOO.Y., Roach, J., Glusman, G., Baliga, N.S., Deutsch, K., Pan, M.,
Dassarma, S., Ng, W.V. and Hood, L.
Low-pass Sequencing for Microbial Comparative Genomics
Unpublished (2003)
Contact: Goo Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoosystemsbiology.org
Seq primer: M13 Forward
Class: shotgun.

FEATURES
source
1. .659
/organism="Haloarcula marismortui"
/mol_type="genomic DNA"
/strain="ATCC 43049"
/db_xref="taxon:2238"
/clone_lib="Hm pUC18 Library"
/notes="Vector: pUC18; Site 1: SmaI; A shotgun library was
constructed from Haloarcula marismortui genomic DNA using
pUC18/SmaI/BAP plasmid"

ORIGIN
Query Match 7.2%; Score 114.2; DB 28; Length 659;
Best Local Similarity 52.2%; Pred. No. 4.4e-07;
Matches 304; Conservative 0; Mismatches 269; Indels 9; Gaps 2;

```

QY 517 GCCACGGACACGGTGGATCTTCCAGCGGTCCGGCGCTGCCGTGCGCGCATGGTC 576
  |||
Db 646 GGCAGCGTGTCCGGCGCGCTCTCTCCGGGATCGGCTGCTGTCGTCGCGCGTGGCGACGCT 587
  |||
QY 577 GACTTCCGCTCGGCGTGCATGAGGAGGAGCAGCATCCGATCCCGCGCGCATGTTGCGGA 636
  |||
Db 586 GACTCGCATCTCGAAGCGCACTGGCGGTTCTATCGGTGCCCGCGCGCTTCTGTGGC 527
  |||
QY 637 CTGCTCGCATGAAGCGACCCACGCGCTGTGTGCCCTTTTACGCGCTGACATACATGGAC 696
  |||
Db 526 GTGCTCGCATCAAGCGGACCTACGGGCTGGTTTCCGGTACGGGCTCATCGGCTACGCT 467
  |||
QY 697 CACACCTTGACACATCGGGCCCATCACAGGGGGTGGAGCTCAACGCCCGGCTCCTC 756
  |||
Db 466 AACAGCTCGAAGATCGGTCCATCGCCCTCTGTTGAGGGCGTGGGAACTGCTG 407
  |||
QY 757 GAGGTGTTCCCGGGCGGACTGCGCGACCCCTCAGTGGGTGCGTAACCTTCGCGAGCGG 816
  |||
Db 406 GATGTCATCGCGGCGCGGACGAAACGACGCGACGACGAGCGAGCGCGCGGCGCGAC 347
  |||
QY 817 GAGAACTACGGCTCCGCGCTCGGCGAGGAGTATCGGCTCGAGATTCGGGTCGTCGAG 876
  |||
Db 346 GGTCTCTACGAGCGGCGCGGACGTTGACGTGACGAGCTCTCTATCGCGCTTCCNACG 287
  |||
QY 877 GAGTCACTGGAGCGGAAACGCTGCGAGCGCGGAGCTGATCGCGCGCTTCAACCGGAGCTG 936
  |||
Db 286 GACTGCTTG-----ACGGGGCTGACGAAGCGTCTGTCGAGAGCTTCTGGGAGCAATG 233
  |||
QY 937 CGGCGCTCGAGAGCGCGGTGCGACCATCGAGGGGTCTCGGTGCGGTGTTGAGAGCGGCG 996
  |||
Db 232 GACGACCTCGAAGCCCGGCGGAGCTACCGAGGTGACCTCCCGTGGTGGAAACAC 173
  |||
QY 997 GCCTGGGCTATCCAGAGCGCGTGGTGGCTTTCAACCGCGCGCTA---TGGGCGACTCC 1053
  |||
Db 172 GCGGTTGAGCGGTACTACGTATCGCATCGGCGCTCTCAACCTCGCGCGGCTTC 113
  |||
QY 1054 GCGGCTGTGGGCTACTTCCCAAGGGCGCGGTGAGAGCTGACG 1095
  |||
Db 112 GACGCGCTCGGCTACGGCCCAATCGGCGGCTACGACGCGCAAC 71
  |||

```

RESULT 8
BZ262005

LOCUS BF262005 857 bp mRNA linear EST 23-OCT-2001
DEFINITION HV_CEA0002M22f Hordeum vulgare seedling green leaf EST library
HVCNDA0004 (Blumeria challenged) Hordeum vulgare subsp. vulgare
cDNA clone HV_CEA0002M22f, mRNA sequence.

ACCESSION BF262005
VERSION BF262005.2 GI:13260083
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Hordeum.
1 (bases 1 to 857)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Nov 17, 2000 this sequence version replaced gi:11192992.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 385
Seq primer: AATTAACCTTCACATAAGGG

High quality sequence stop: 509.

FEATURES
source

Location/Qualifiers
1. .857
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="C116155 (M1a13)"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HV_CEA0002M2f"
/tissue_type="seedling green leaf"
/lab_host="TJC121"

/clone_lib="Hordeum vulgare seedling green leaf EST
library HVCDNA0004 (Blumeria challenged)"
/note="vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
C.I. 16155 (M1a13) plants were greenhouse grown in the R
wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were challenged with isolate A27
(AvrM1a13) of Blumeria graminis f. sp. hordei, and leaves
were harvested 20 and 24 hr post-inoculation and snap
frozen; uninoculated leaves were harvested 20 hr
post-inoculation (Wei, Wise). In the TJ Close lab at the
University of California, Riverside, total RNA was
prepared from each sample pool, equal quantities of all
three RNA pools were combined, poly(A) RNA was purified
from the mixture, one cDNA library was made, and 1 million
pfu were in vivo excised to give pBluescript SK(-) cDNA
phagemids (Choi, Close). Phagemids were plated and picked
at the Clemson University Genomics Institute (CUGI)
(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA
preparations, DNA sequencing and sequence analysis were
performed at CUGI (Wing, Yu, Frisch, Henry, Simmons,
Oates, Rambo, Main). The sequence has been trimmed to
remove vector sequence and contains a minimum of 100 bases
of phred value 20 or above. For more details on library
preparation and sequence analysis see
http://www.genome.clemson.edu/projects/barley. To order
this clone see http://www.genome.clemson.edu/orders Also
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
Genetically and physically anchored EST resources for
barley Genomics. Barley Genetics Newsletter 31:29-30.
(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html")

ORIGIN

Query Match 7.2%; Score 113.4; DB 10; Length 857;

Best Local Similarity 55.3%; Pred. No. 6.2e-07;

Matches 286; Conservative 0; Mismatches 221; Indels 10; Gaps 3;

QY 262 CGGTGGAGGCGGCACGAGGGCCACTGAGCACTGACCGCGCGATCAGGACTGC 321
DB 91 CAGGAGCGTACCCCTGCAGGCTCACCCTTCGCCATCAGGACATCTTCTAGTCGCC 150
QY 322 ATGCCCATTCGGGTATGCCACCAAGCGGTCCCGATGCTCCGACTGTGATCGCC 381
DB 151 GGCGCGTCAAGGGTTCGGACCCCGGACTGGGCGCGGACGACGCGCGCGCCACC 210
QY 382 ACCGAGGATCCGTGGTGGAGGGGTGCTCCGCGGACGGCGGACCATCTGTCGGAG 441
DB 211 ACCGCGCCCGCTGCTGGCGC-----GCTCGCGCGGCGGCGGACGCGGCGTGGCATG 264
QY 442 ACGAACCTCGAGGACATGGCGATGGGTAT---CGGTGAAGGCGACGCTACGGTCTTCGCG 498
DB 265 ACGGTGATGGACGAGATGGCTACAGCATCAACGCGGAGACGCGCATACGGCACCCCC 324
QY 499 CTGAACCCGACACACCCCGCCACGGGACGGGTGGATCTTCACGGGCTCGGGCGTGGC 558
DB 325 GCCAACCCCTGGCGCCCGCGGGTCCCGCGCGGTTCCTCCACGCGGCTCGCGCGTGGC 384
QY 559 GTCGGTGCGGCGATGGTGCATCTTCGCCCTGGCGGTTCGATGAGGAGGCGAGCATCCGGATC 618
DB 395 GTGACGCGACGCTGCGCGCATCTTCGCCCTCGGACCGGACACCGGCGGACGGTCAGGGTG 444
QY 619 CCGCGCGCATGTTGGG-GACTGTGTCGATGAAGCGGACCCACCGGCTGTGTGCGTCTTA 677

DB 445 CCGCGCGCATACTGCGAGCATCTTCGGCTTACGCCCTCCACGCGAGGGCTCCGCGGA 504
QY 678 CGGCTCTGACATATGACACACACCTTGTGACACATCGGGCCCATCACAGGGGGTCTGA 737
DB 505 GAACGTCTTCCCTCGGAGCAGATGTTTCGACACTCTCGATGGTTAGCCAGAGATGTAGC 564
QY 738 GCTCAACGCCCGGGTCTTCGAGGTTGTTGGCGGGGCC 774
DB 565 TACATTGTCCCGTGTGAGCTACGTGTTACCTGCCGC 601

RESULT 9
LOCUS

CD874694

DEFINITION

AZO3.102N04F011001 AZO3

mRNA sequence.

ACCESSION

CD874694

VERSION

CD874694.1

GI:32558510

KEYWORDS

EST.

SOURCE

ORGANISM

Triticum aestivum (bread wheat)

Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Triticum.

REFERENCE

1 (bases 1 to 535)

Genoplante, a major partnership french program in plant genomics

Unpublished (2003)

Genoplante

Contact: Genoplante

93, rue Henri Rochefort 91025 EVRY CEDEX France

Tel: 33 1 69 47 54 00

Fax: 33 1 69 47 54 10

This sequence has been generated in the framework of the french

plant genomics programme 'Genoplante' (http://www.genoplante.com

and http://genoplante-info.infobiogen.fr).

Location/Qualifiers

1. .535

/organism="Triticum aestivum"

/mol_type="mRNA"

/cultivar="recital"

/db_xref="taxon:4565"

/clone="AZO3102N04"

/tissue_type="leaf"

/clone_lib="AZO3"

ORIGIN

Query Match 7.1%; Score 112.4; DB 14; Length 535;
Best Local Similarity 60.1%; Pred. No. 7.7e-07;
Matches 226; Conservative 0; Mismatches 141; Indels 9; Gaps 2;
QY 279 GGAGGGGCGACATGAGCGACCTGACCGCGCGATCAAGGACTGCATCGCCATGCGCGPAT 338
DB 138 GCAGCTACCCCTGCACGGCTCACCTTCGCTATCAAGGACATCTTCGACGTCGCGGCGC 197
QY 339 GCCCACCAGAACCGGTGCC-----GGATGTCCTCCGACTGTCGTCGACCGAGGATGC 392
DB 198 CBTACCGGCTTCGGACCCCGGACTGGGCCCGGACGACGCGCGCGCGCGCGCAACCGC 257
QY 393 CBTGTTGTTGAGCGGCTGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 452
DB 258 CCGCGCGTTCCTGGCGCGGCTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 317
QY 453 GGACATGGCGATGGGTAT---CGGTGAAGGCGGCTTACGGTCTCTGCGTGAACCCGNA 509
DB 318 CGAGATGGGCTTACAGCATCAACGGCGAGAAACGCGACTACGGCACCCCGCGCAACCCCTG 377
QY 510 CAAACCCCGCGCGCGGCGGCTCTTCCAGCGGCTCCGCGCGTCCGCGTCTGCGTCCGCGG 569
DB 378 CGCCCTGCGCGGCTCCCGCGGCTCTCTCCAGCGGCTCCGCGCTCGCGCTCGCGCGACG 437
QY 570 CATGTGCGATCTTCGGCTTGGCGCTGATGAGCGAGGACGATCCCGATCCCGCGCGCATG 629

Db 438 CTTCCGCGACTTCGCCCTCGGACCGACCGGCGGAGCGTCAGGTTGCCCGCGCCTA 497

QY 630 GTGCGGACTGTTCCGC 645

Db 498 CTGCGGCACTTTCGCG 513

RESULT 10

AZ935507 640 bp DNA linear GSS 24-APR-2001

LOCUS BJ_Ba0003L22f B. japonicum BAC library Bradyrhizobium japonicum

DEFINITION genomic, genomic survey sequence.

ACCESSION AZ935507

VERSION AZ935507.1 GI:13778331

KEYWORDS GSS

SOURCE Bradyrhizobium japonicum

ORGANISM Bradyrhizobium japonicum

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE 1 (bases 1 to 640)

AUTHORS Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A., Goicoechea, J.L., Stacey, G., Sadowsky, M.J., and Wing, R.A.

TITLE A marker-dense, sequence-ready map of the Bradyrhizobium japonicum Genome

JOURNAL Genome Res. 11 (8), 1434-1440 (2001)

MEDLINE 21376150

PUBMED 11483585

COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Class: BAC ends
High quality sequence stop: 582.

FEATURES

source Location/Qualifiers

1..640

/organism="Bradyrhizobium japonicum"

/mol_type="genomic DNA"

/strain="USD110"

/db_xref="taxon:375"

/lab_host="E. coli"

/clone_lib="B. japonicum BAC library"

/notes="vector: pindigo536; Site_1: HindIII"

ORIGIN

Query Match 7.1%; Score 111.4; DB 28; Length 640;

Best Local Similarity 58.3%; Pred. No. 1.1e-06;

Matches 215; Conservative 0; Mismatches 151; Indels 3; Gaps 1;

QY 257 TCTGCCGGTGGAGGGCGCACGGAGGGCCACTGAGCGACTGACCGCGCGATCAAGG 316

Db 369 TCTGCGATGCGGCGCGCGCGCGATCGGACCGCTGATGCGGTGCGGATCAAGG 310

QY 317 ACTGCAATCGCATCGCGGTATCCCAACAGCGGTCCCGGATGCTCCCGACTGTGA 376

Db 309 ACTCACCGATACCGGGGCTGACCAACCACTACGGCTCGGCTTGTTCGGATCAG 250

QY 377 TCGCCACCGAGGATGCGGTGGTGGAGCGGTGCTCGCGGAGCGGCCACCATCGTGC 436

Db 249 TCCCGGCCGAGGACGAGCTGTTGTCGCGCGGTGCGGCGCGCGCGGATCATTTGG 190

QY 437 GCAGAGCACTCGAGGACATGCGCATGGT---ATCGGTGAGGACCGCTACGGTC 493

Db 189 GAAAGACCAACACCGCGGAATTCGGTTCGGTCCGGTCTGCAACACCGGTTCGCGGGC 130

QY 494 CTGCGGTGAACCGCAACACCGCGCCACCGCACCGGTGATCTTCACGCGGTCCGGCG 553

Db 129 CGAGCGCAATCTCTTCGATCCCGCGCTGACCTCCGGCGGATGCTCCGCGGCTCTGCG 70

QY 554 CTGCGGTGCTGCGGCACTTCGATCTTCGCCCTGGCGGTGATGAGGAGGACGATCC 613

Db 69 TTCCGGTCCGCGCGCATGTTAACGCTGGCGGACCGACCGATTTTCGGGGCTCGGTGC 10

QY 614 GGATCCCGG 622

Db 9 GCACGCCCG 1

RESULT 11

AZ935188 604 bp DNA linear GSS 24-APR-2001

LOCUS BJ_Ba0003B03f B. japonicum BAC library Bradyrhizobium japonicum

DEFINITION genomic, genomic survey sequence.

ACCESSION AZ935188

VERSION AZ935188.1 GI:13777512

KEYWORDS GSS

SOURCE Bradyrhizobium japonicum

ORGANISM Bradyrhizobium japonicum

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE 1 (bases 1 to 604)

AUTHORS Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A., Goicoechea, J.L., Stacey, G., Sadowsky, M.J., and Wing, R.A.

TITLE A marker-dense, sequence-ready map of the Bradyrhizobium japonicum Genome

JOURNAL Genome Res. 11 (8), 1434-1440 (2001)

MEDLINE 21376150

PUBMED 11483585

COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Class: BAC ends
High quality sequence stop: 553.

FEATURES

source Location/Qualifiers

1..604

/organism="Bradyrhizobium japonicum"

/mol_type="genomic DNA"

/strain="USD110"

/db_xref="taxon:375"

/lab_host="E. coli"

/clone_lib="B. japonicum BAC library"

/notes="vector: pindigo536; Site_1: HindIII"

ORIGIN

Query Match 6.8%; Score 107.6; DB 28; Length 604;

Best Local Similarity 53.1%; Pred. No. 4e-06;

Matches 277; Conservative 0; Mismatches 239; Indels 6; Gaps 2;

QY 270 GGGCGCCACGGAGGGCCACTGAGCGACTGACCGCGCGATCAAGGACTGCATCGCAT 329

Db 11 GGATGCCCGCAGCATGCGGCTCTATGCGGTCCCGGTGAGGACAAACATCGACGC 70

QY 330 CGCGGTATGCCACCAACGAGCGGTCCCGAGTCTCCGACTGTGATCGCACCGAGGA 389

Db 71 GCTGGGCTTTCCGACACG---GCGGCTGCCCGCCTTCTCTACACCGACCCATGA 127

QY 390 TCCCGTGGTGGAGCGGCTGCTCGCGGAGGGCCACCATGCTCGGCAAGACGACCT 449

Db 128 CTCGACCGCGTGGAGGCGCTGCGCGGCGCGCGCATCATCATCGGCAAGACCAATCT 187

QY 450 CGAGGACATGCGCATGGG---TATCGGTGAAGGCGAGGTCTACGGTCTCGCTGAACCC 506

Db 188 CGACCATGTCGAACCGCGCTCTGTCGGGTACGTCGCCCTACGGCATTCGCAAAATTC 247

QY 507 GAACAAACCCGCCCAACCGGACGCGGTGATCTTCAGCGGTTCGCGCGCTCGCTGCTGC 566

Db 248 GATTGCGAGGATCTGATTCGCGGTGGCTTCAGCTCGGGGTTCGCGGACCGCCCTCGGCG 307

QY 567 CGGCATGTCGACTTCGCCCTGGCGGTGATGAGGAGGAGCATCCGGATCCCGGCGGC 626

```

Db 308 CGGGCTGTCGGCTGACCTCGGCACCGACACCGCGCGAGCGCGCTGCGGGCCAT 367
Qy 627 ATGTGCGGACTGTCGGCATGAGGCGACCCAGCGCTGGTCCGTCATTACGGCTGAC 686
Db 368 GCTCAACAACATCTCGGCTGGAAGCAGCCTCGGCATGATCTCGACCGCGGACTCGT 427
Qy 687 ATACATGAGACACACCTTTGACCATCATCGGCCCATCACCAGGGGGTTCAGTCAACGC 746
Db 428 GCCTGCTGCGGACGCTGACTCATCTCGGTATTGCGCTTGACGGTGGACGATGCTGC 487
Qy 747 CCGGGTCTCGAGTGTGGCGGGCGGACTGCGCGGACCC 788
Db 488 ACTGCACTTCTGTGATGGCGGACCTGACCAAGCCGATCC 529

RESULT 12
LOCUS AQ012146/c
DEFINITION 875011B093011698 Cosmid library of chromosome II Rhodobacter
sphaeroides genomic clone 875011B093011698, genomic survey
sequence.
ACCESSION AQ012146
VERSION AQ012146.1 GI:3177101
KEYWORDS GSS.
SOURCE Rhodobacter sphaeroides
ORGANISM Rhodobacter sphaeroides
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
Rhodobacteraceae; Rhodobacter.
REFERENCE 1 (bases 1 to 2106)
AUTHORS Choudhary,M., Mackenzie,C., Mouncey,N., Weinstein,G.M. and
Kaplan,S.
TITLE RsgDB, the Rhodobacter sphaeroides Genome Database
JOURNAL Unpublished (1998)
COMMENT Contact: Choudhary, M.
Department of Microbiology and Molecular Genetics
University of Texas Medical School
6431 Fannin Street, Houston, TX 77030, USA
Tel: 713 500 5437
Fax: 713 500 5499
Email: madhu@utmsi.med.utmc.edu
Seq primer: pBluescript T3
Class: shotgun.

FEATURES
source
1. .2106
/organism="Rhodobacter sphaeroides"
/mol_type="genomic DNA"
/strain="2.4.1T"
/db_xref="taxon:1063"
/clone="875011B093011698"
/lab_host="E. coli S17-1"
/clone_lib="Cosmid library of chromosome II"
/note="Vector: pLA2917"

ORIGIN
Query Match 6.8%; Score 106.6; DB 28; Length 2106;
Best Local Similarity 52.3%; Pred. No. 7.6e-06;
Matches 285; Conservative 0; Mismatches 254; Indels 6; Gaps 2;

Qy 295 GACATGACCGCGCGGATCAGGACTGTCGCCATCGCGGTATGCCACACGACGCGG 354
Db 895 GGCATCCCTCGCGCTGAGGACGAGACCATGATGCGGGCAGGTTCAGACCTACGCG 836
Qy 355 TCCGGATGCTCCCGACTGTGATGCCACCGAGGATCGCGTGTGGTGGAGCGGCTGCTC 414
Db 835 TCGTCTTTTACGCGGACCATGTGCTGCGAAGACGCGCGGTGCTGCGCGCTGGTC 776
Qy 415 GCGGAGCGCGCACCATC ---GTGGCAAGACGACCTCGAGGACATGGCGATGGGATC 471
Db 775 GAGGCGGCGCGGTGATCATCGCCGACACACGACCGCCGATTCCTCTGGCTCCCTTC 716
Qy 472 GGTGAGCGACGCTTACGCTCTGCTGATGAACCCGAAACACCCCGCCACGCGCGGT 531
Db 715 TGTCACTCGCGGCACTGGGGGCTGACGCGCAACCCGTAACCTTCGCGCTTCAAGCCGCGC 656

```

```

Qy 532 GATCTTCCAGGGCTCGGGCTGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGGC 591
Db 655 GCTCTCTCGGGCGCTCGGGCGCGCGCTCGGGCGGGCTCGGGCGCGCTCGGGCGCGC 596
Qy 592 GTGATGAGGAGGAGGAGGATCCGGATCCCGCGCGCATGGTGGGACTGGTGGCATGAAG 651
Db 595 TCGGACATCGGGCGCTCGATTGGATCCCGGCTCGGCTCGGGGTGGTGGGTTCAG 536
Qy 652 GCGACCCAGCGCTGGTGGCTCTTACGGCTCGACATACATGACGACCACTTGGACCC 711
Db 535 CCGCCCTACGGCGGTGCTGCCCTCGAC ---CCGCGCTTCAACCTCGATCCCTACAAC 479
Qy 712 ATCGGGCCCATCACGAGGGGGTTCGAGCTCAACCGCGGCTCGAGGTGTCGAGGTGTCG 771
Db 478 CCGGGCCCATCGGCGCGCATGCGTGAAGATTGCTGATGTCAGAACGTCGCTCGCGGC 419
Qy 772 GCGGACTGCGGCGACCTCAGTGGTGGTGAACCTTCGGGAGCGGAGAACTACGGCTCC 831
Db 418 CCCCATCGGAGACATCGCTCGCTGAAACCCAACTCGAACTCGCGCTCGACCGGTGCG 359
Qy 832 GCGCT 836
Db 358 GGGT 354

RESULT 13
LOCUS AZ934731
DEFINITION AZ934731 521 bp DNA linear GSS 24-APR-2001
genomic, genomic survey sequence.
ACCESSION AZ934731
VERSION AZ934731.1 GI:13776791
KEYWORDS GSS.
SOURCE Bradyrhizobium japonicum
ORGANISM Bradyrhizobium japonicum
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobaceae; Bradyrhizobium.
REFERENCE 1 (bases 1 to 521)
AUTHORS Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A.,
Goicoechea,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
genome
Genome Res. 11 (8), 1434-1440 (2001)
MEDLINE 21376150
PUBMED 11483585
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Class: BAC ends
High quality sequence stop: 482.
FEATURES
source
1. .521
/organism="Bradyrhizobium japonicum"
/mol_type="genomic DNA"
/strain="USD110"
/db_xref="taxon:375"
/lab_host="E. coli"
/clone_lib="B. japonicum BAC library"
/note="Vector: pIndigo536; Site 1: HindIII"

ORIGIN
Query Match 6.7%; Score 106.2; DB 28; Length 521;
Best Local Similarity 53.2%; Pred. No. 6.1e-06;
Matches 272; Conservative 0; Mismatches 233; Indels 6; Gaps 2;

```

```

Qy 270 GCGCGCCAGGAGGCGGCGGACTGAGGACTGACCGCGGATCAAGGACTGATCGCAT 329
Db 11 GGATGCGCGCAGCATCGCGCTCTATGGGTCCCGGTCCGGTGAAGGACCAATCGAGC 70

```

```

email: rzwang@cs.cmu.edu
Total hq bases = 213
Seq primer: AATTACCCCTCACTAAAGG
High quality sequence stop: 438.
Location/Qualifiers
1. .517
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2004, 04:34:02 ; Search time 125.584 Seconds
(without alignments)
6918.029 Million cell updates/sec

Title: US-10-658-691-4
Perfect score: 1578
Sequence: 1 ggcgacccaatgcgccatt.....ggcaccgggtgcggtag 1578

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	224.8	14.2	1521	1 US-08-726-136-20	Sequence 20, Appl
2	224.8	14.2	1521	3 US-03-103-434-20	Sequence 20, Appl
3	224.8	14.2	1521	3 US-08-687-594-20	Sequence 20, Appl
4	180.8	11.5	1879	1 US-08-539-866-1	Sequence 1, Appl
5	172.8	11.0	4403765	3 US-09-103-840A-2	Sequence 2, Appl
6	170.4	10.8	4411529	3 US-09-103-840A-1	Sequence 1, Appl
7	168.8	10.7	1878	1 US-07-612-673-1	Sequence 10, Appl
8	168.4	10.7	1879	1 US-07-796-361A-10	Sequence 10, Appl
9	151	9.6	1467	4 US-09-252-991A-9752	Sequence 10084, A
10	151	9.6	1818	4 US-09-252-991A-10084	Sequence 2, Appl
11	123.4	7.8	4403765	3 US-09-103-840A-2	Sequence 1, Appl
12	123.4	7.8	4411529	3 US-09-103-840A-1	Sequence 8, Appl
13	120	7.6	2456	1 US-08-233-146-8	Sequence 8, Appl
14	120	7.6	2456	1 US-08-463-470-8	Sequence 8, Appl
15	114.6	7.3	1335	4 US-09-252-991A-15844	Sequence 15844, A
16	114.6	7.3	1746	4 US-09-252-991A-15790	Sequence 15790, A
17	114.6	7.3	3036	4 US-09-252-991A-15848	Sequence 15848, A
18	110.4	7.0	23673	4 US-09-773-816-1	Sequence 1, Appl
19	108.8	6.9	1843	1 US-07-948-023-1	Sequence 1, Appl
20	104.2	6.6	1209	4 US-09-252-991A-9930	Sequence 9930, Ap
21	102.2	6.5	477	4 US-08-252-991A-15815	Sequence 15815, A
22	100	6.3	1401	4 US-09-489-039A-2359	Sequence 2359, A
23	93.8	5.9	1932	4 US-09-252-991A-10046	Sequence 10046, A
24	91.8	5.8	1395	4 US-09-252-991A-12734	Sequence 12734, A
25	91.8	5.8	1434	4 US-08-252-991A-12585	Sequence 12585, A
26	90.8	5.8	1462	3 US-08-961-083-41	Sequence 41, Appl
27	90.8	5.8	1462	4 US-09-536-784-41	Sequence 41, Appl

Query Match 14.2%; Score 224.8; DB 1; Length 1521;

28 90.8 5.8 1476 4 US-09-252-991A-7653 Sequence 7653, Ap
29 90.8 5.8 1485 4 US-09-252-991A-7738 Sequence 7738, Ap
30 90.8 5.8 2997 4 US-09-252-991A-7831 Sequence 7831, Ap
31 90.8 5.8 19702 4 US-08-961-527-7 Sequence 7, Appl
32 88.4 5.6 1401 1 US-08-726-136-27 Sequence 27, Appl
33 88.4 5.6 1401 3 US-09-103-434-27 Sequence 27, Appl
34 88.4 5.6 1401 3 US-08-687-594-27 Sequence 27, Appl
35 87.4 5.5 1640 1 US-07-796-361A-12 Sequence 12, Appl
36 87.4 5.5 1816 1 US-08-539-866-3 Sequence 3, Appl
37 87.4 5.5 1817 1 US-07-612-673-3 Sequence 3, Appl
38 80.4 5.1 44377 2 US-08-804-227C-7 Sequence 7, Appl
39 80.4 5.1 44377 2 US-08-804-198-1 Sequence 1, Appl
40 78.4 5.0 1230025 4 US-09-198-452A-1 Sequence 1, Appl
41 76 4.9 876 4 US-09-134-000C-1186 Sequence 1186, Ap
42 68.6 4.3 1476 3 US-08-896-346-1 Sequence 1, Appl
43 67.2 4.3 1347 4 US-09-107-532A-308 Sequence 308, App
44 66.2 4.2 2045 3 US-08-743-168B-42 Sequence 42, Appl
45 66.2 4.2 2045 3 US-08-743-168B-44 Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-08-726-136-20
; Sequence 20, Application US/08726136
; Patent No. 5811286
; GENERAL INFORMATION:
; APPLICANT: ROBERT D. FALLON
; APPLICANT: MARK S. PAYNE
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING
; TITLE OF INVENTION: STEREOSPECIFIC NITRILE HYDRATASE AND AMIDASE ENZYMES AND
; TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR
; TITLE OF INVENTION: THE PRODUCTION OF CHIRAL AMIDES AND ACIDS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
; SOFTWARE: MICROSOFT WORD 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,136
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/004914
; FILING DATE: OCTOBER 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA A.
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1521 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
US-08-726-136-20

Best Local Similarity 48.8%; Pred. No. 7,8e-35;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;

QY 37 CGGACAGCCGACAGCTCCAGAGGTACAGGCGCCGACCACTTCGACCTCGACGAGGAA 96
DB 16 CTTACCTCGACAGGTTTACATCATCGAACCCAGTTGACATGCACTGACGACGAA 75
QY 97 GTGGCGGCCAGCTGTTCCGGTCTGTGGGAGATGTTGACTGCTTCGACCTGATCGAC 156
DB 76 CAGGAGAGTCTACTGGAAGTATGAAACGAGTTTCGAGCTTACGACCTGATCGAC 135
QY 157 GAACTACCGCAACCCGCGAGCCGCGGACCGGTTACAGCAACGCGATCGGCGCGGAA 216
DB 136 GAACTGGGTGATTTGGTTCGCGCAATACGCTACGACCGCAGTTGAGGTATGCGCATCGG 195
QY 217 CGGACCGGAGAGAGACCGTTCAACGATTCATCCGTTTCGCGGGTGGAGCGGCC 276
DB 196 CCACTCGGCAAGAAACCTCTGAACCGCTGTGTACTACGAAACAGAAAGTAAATGGTGCC 255
QY 277 ACGGAGGCGCACTGAGCGACCTGACCGCGCGATCAAGGACTGATCGCCATCGCCGT 336
DB 256 CGCGAAGGCTCTGCGCGGCAAAACCGTTCGCGCTCAAAGATAATATCTCCTCGCAGGC 315
QY 337 ATGCCCAACCAAGCGGTCCCGATGCTCCGACTGTGATCGCCACCGAGATGCGGTG 396
DB 316 GTCCCATGATGAACGGCGAGCGCGCTTGAAGGCTTCGTCGCGGGTTGATGCCACG 375
QY 397 GTGTGGAGCGCTGCTCGCGCAGCGCCACCATCGTCGCGCAAGACCACTCGAGGAC 456
DB 376 GTGGTCAACCGCTTGTCTGATCGCGGGCGACCAATCTCGCAAGACCACTCGGAGCAC 435
QY 457 ATGGCGAT---GGGTATCGGTGAAGGAGCGTCTAGGTCTGCGTGAACCGCAACAC 513
DB 436 TACTGCTTTTGAAGGAGGAGCGACACACTCCGATCCAGCGCGCGGTGACAAACCAATCGC 495
QY 514 CCGCGCCACGCGAGCGGTGGATTTCTCAGCGCGCTCCGCGTTCGCTGCGCGGATG 573
DB 496 CACGGTTATGCTCTGGCGTTCTCATCAGCAGCGCGCATTTGTTGCGTCCGGTAG 555
QY 574 GTGACTTCGCTCGGCGTTCATGAGGAGCAGCATCCGATCCCGCGCGCATGTCG 633
DB 556 GTGGACATCGCGTGGCGGCGATCAAGCGGCTCCATTCGATCCCGTCCGCGCTTCG 615
QY 634 GAGCTGTGCGATGAAGCGACCAACCGCGCTGTGCGCTTACGCGCTGACATACATG 693
DB 616 GGTACCTACGATGAAGCCACCCACCGCTGTGCTTACCGCGTTCATGCGGAT 675
QY 694 GACCAACCTTGAACACATCGGCGCCATCACAGGGGTTGAGCTCAACCGCGGTC 753
DB 676 GAAGCCACGATGATCATGTCGCGCCCATCACCGGTAAAGTGGCGGCAAGCGCTGATG 735
QY 754 CTGAGGTGTGGCGGGCGGACTGCGCGGACCTCATGTGGTGGTAACTTCGCGAG 813
DB 736 CTGAGGCAATGGCGGTGACAGGACTGACCGCGCGCGCGGCGG-----CTCAG 789
QY 814 CCGGAGAACTAGCTTCGCGTTCGCGGAGGAGTATCGGTTCTGAGTTCGCGTTCGTC 873
DB 790 GTGATGACTATTCAGTTTACCTGGAAGAGCGGTGAGCGGACTCAGAATCGGGTGTG 849
QY 874 GAGGAGTCACTGAGCGCAACCGTGCAGCGCGGCTGATCGCGCGTTCACCCAGGA 933
DB 850 CHAGAGGATTCGCTTCTAACAGGACCTTCGCGTGGCGGCAAAAGTGGCGACGCC 909
QY 934 CTGGCGGCTCTGAGAGCGCGGTGCGACCATCGAGCGGTTCTCGGTGCGGTTGTGGACG 993
DB 910 ATCGCGGACTCGAGCGGTGGGCGTCTATGTGAGCGGCTCTCCATTCGCGAGCACAAC 969
QY 994 CGGCTGCGCTATCAGAGCGCGGTGATGGCTTTCAACGCGCGGCTATGCGGACTCC 1053
DB 970 CTGGCAGGTTGTGTGGCACCCCATTCGCTTGGAGGCTTGACCATGAGATGAT 1029
QY 1054 GCGGTGTGGGTACTTTCCACAGGGCGCGTGGAGCTCAGCACCGCGCTCACGACGCG 1113

1030 GGCAACGGCGCAGGCTTTAAGTGGAAAGGACTTTAGATGTGCGGCTGCTGGCAACAA 1089
1114 GCCAGAGTCGACACCAACCAAGGATCTGGCGATCTCTCCGGGTGATGCTGTGATC 1173
1090 GCCAGTGGCGGACGACGACCAATATATCCGCTGCTCAAGCTCTGATGTTGTC 1149
1174 GCGGACACCTGCGGACGAATACCTCGGCATCTCACTAGCGAAGCGCAGAACCTGCGG 1233
1150 GGCAATAAGGCTGTGCGCTACACGACGCTACTAGCCCAAGGCCCAAGCTTGCA 1209
1234 CTGGAGCTCGCAAGCAGATCGAGCGCTTCCTCAGGACCGGGCTGCACTGTGACCCG 1293
1210 CGCTTTCGCGGAGGATACGACAAAGCGCTGCAAACTATGACCTGCTGCTGATCGG 1269
1294 ACCAGCCTACGTTCCCAACGAGCTGTTGAGCGGTGCGCAAGACCACTATGCTCATGTC 1353
1270 ACCAGCCCATACGCGCCCAACCCCAACCGCCAG---CGAATGCTGCTGATCAGGAGTAC 1326
1354 CCAGGATGACGGGCAATCGATCTCAACAGTTCGCGCTGCGCTGCGCTTCAAGTGATA 1473
1327 GTGGCTCGCGGCTTGAATATGATCGCAATACCGCGCCACAGGACATCACCGGCAATCG 1386
1414 GCCTGACGCTGCCACGCGTGGCGGCGAGAGGCGCTGCGCTTGGCTTCAAGTGATA 1473
1387 GCAATGCTGATC---CGTGGCTGCTGAGCGCTGCGCTGCGGCTGATGCTGCTGTC 1443
1474 GGCGGCACTTCGAGGAGTCAAGCTTACCGACCGCGCGCTGATGAGGCGCGCG 1531
1444 GCAAAACACTACGCGGAGGCGAGATTACCAAGCGCGCGCGCTTGAAGCGCTCGG 1501

RESULT 2
US-09-103-434-20
; Sequence 20, Application US/09103434
; Patent No. 6133421
; GENERAL INFORMATION:
; APPLICANT: ROBERT D. FALLON
; APPLICANT: MARK S. PAYNE
; APPLICANT: MARK J. NELSON
; TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING
; TITLE OF INVENTION: STEREOSPECIFIC NITRILE HYDROLASE AND AMIDASE ENZYMES AND
; TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR
; TITLE OF INVENTION: THE PRODUCTION OF CHIRAL AMIDES AND ACIDS
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
; SOFTWARE: MICROSOFT WORD 2.0C
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,434
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,136
; FILING DATE:
; APPLICATION NUMBER: 60/004914
; FILING DATE: OCTOBER 6, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA A.
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CR-9677
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:
LENGTH: 1521 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
US-09-103-434-20

Query Match 14.2%; Score 224.8; DB 3; Length 1521;
Best Local Similarity 48.8%; Pred. No. 7.8e-35;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;
QY 37 CCACAGCCGACAGCTCCAGAGGTACAGCGCCGCCACACCTTCGACCTCCAGAGGAA 96
Db 16 CTTACCTCGACCCAGGTTTATAGACATCCGAACCCAGTTTCGACATGCAACTGACGACGAA 75
QY 97 CTGGCCGCCAGCTCGTTCCGGTCGTCGGGAGATGTCGACTGCTTCGACCTGATCGAC 156
Db 76 CAGGAGAGCTCTACTGGAACCTGATGCAACCGAGTTTCGACGCTTACGACCTGTCGAC 135
QY 157 GAACTACCCGAACCCGCCGAGCGCCGACGCCGTACACGGAACGGGACATCGCCGGA 216
Db 136 GAACTGGCTGATTCGTTCCGCCAATACGCTACGACCGCAGTTTCAGGCTATCGCCATCG 195
QY 217 CCGACCGGCGACAGACCGGTTCAACCGATTTCATCCGGTTCTGCGGGTGGAGGCGCC 276
Db 196 CCATCGCCGACGAAACCTCTTGAACGCTTGTACTACCGAACAGAGTGAATGTGCC 255
QY 277 ACGGAGGGGCACTGAGCGACCTGACCGCGCGATCAAGGACTGCATCCCATCGCCGT 336
Db 256 CGCGAAGGCTGTGCGGGGCAAAACCGTCGCGCTCAAGATAATATCTCCCTGGCAGCG 315
QY 337 ATGCCACACGAAACCGGTCGCGATGCTCCCGACTGTGATGCCACCGAGGATCCCGT 396
Db 316 GTCCCATGATGAACCGCGCAGCGCGCTTGAAGGCTTGTCCCGGGTTCGATGCCACG 375
QY 397 GTGGTGGAGCGGCTGTGCGGGCAGCGCCACCATCGTCGGAAGACGAACTTCGAGGAC 456
Db 376 GTGGTACCCGCTTGTCTGATCGGGGGGCAACATCTCGGCAAGCCACCTCGGAGCAC 435
QY 457 ATGGCAT---GGTATCGTGAAGGAGCGCTTACGCTGCTCGCTGAACCGCAACAC 513
Db 436 TACTGCTTTTCAGGAGGACCCACACCTCGCATCCAGCCCGGTGCACACCCACATCGC 495
QY 514 CCGCCCAACGCGAGCGGTGATCTTCAGCGGCTCGGGGCTCGCGTCCGTCGCGGCATG 573
Db 496 CACGGTTATGCTCTGCGGTTCTCTATCAGGAGCGCGCATGTTGGTCCGCTGAG 555
QY 574 GTGACTTCGCTTGGCGCTGATGAGGAGGAGCATCCGATCCGCGCCGATGGTGC 633
Db 556 GTGACATCGCGTGGCGCGATCAAGCGGCTCCATTCCGATCCGCTCGCGCTTCTGC 615
QY 634 GGAATGCTGCGATGAAGGCGACCCAGCGCTGTCGCTTACGCGCTGACATACATG 693
Db 616 GGTACCTAGGATGAAGCCACCCAGCGCTGTGCTTACACCGCGCTCATGGCGATT 675
QY 694 GACACACCTTGGACACATCGGCGCCATCACCAGGGGGTTCGAGCTCAACCGCGCGTC 753
Db 676 GAAGCCAGATCATGATGTCGCGCCCATCACCAGTAACTGTCGCGCAACCGCTGATG 735
QY 754 CTCGAGTGTTCGCGGGCGGCTGCGGACCTCAGTGGGTGCTTAACCTTCCGGAG 813
Db 736 CTGAGGCAATGGCGGTGCGAGCGGATCGTACCCCGCGCAGGGCGG-----CTTCAG 789
QY 814 CCGGAGAACTACGGCTCGCGCTCGGAGGAGTATCCGCTCTGAGATTCGGGGTGC 873
Db 790 GTCGATGACTATTGAGTTACTCTGAAAGAGCGCTGAGCGGACTCAGAACTCGGGTGTG 849
QY 874 GAGAGTCACTGGAGCGGAACGGTTCGACCGCGGACGCTGATCGCGGCTTCAACAGGGA 933
Db 850 CAAGAGGATTCGCGCTTGTCAACAGGACCTTCGCTGGCGGCAAAAGTGGCGACGCC 909

QY 934 CTGGCGCGCTCGAGAGCGCGGCTGTGCAACCATCGAGCGGCTCTCGGTGCGTGTGAGCG 993
Db 910 ATGCCCGACTCGAGGCGTGGCGCTCATGTGAGCGCGTCTCAATTCGAGCACAAC 969
QY 994 GCGGCTGGGCTATCCAGAGCGCGGCTGATGGCTTCAACGCGCGCGCTATGCGGACTCC 1053
Db 970 CTGGCAGGGTGTGTGGCAACCCATCGGTTGGAGGCTTGACCATGAGATGATCAT 1029
QY 1054 GCGGCTGGGCTACTTCCACAGGGGCGGCTGTGAGAGCTCAGCACCGCGCTCAAGAGCGG 1113
Db 1030 GGCAAGCGCGCAGGCTTTAACTGGAAGGACTTTACGATGTCGCGCTCTGTCGACAAAC 1089
QY 1114 GCCCAGAGTCCGACCCACACAGGATCTCGCATCTCTGCTCCGCGCTGATGCTGCTGATC 1173
Db 1090 GCCAGCTGGCGGACGACGAGCAATATTCGCGCTCTCAAGCTCTGCTGATGTTGCTC 1149
QY 1174 GCGGAGCACCTGCGCGACGAATACCTCGCATCCACTACGCGAGGCGCGAGAACCTCGCG 1233
Db 1150 GGCCAAATAGGCGCTGTGCGCTACAAAGGAGCTTACTACGCAAGGCCAGAACCTTGCA 1209
QY 1234 CTGGAGCTGGCAAGAGATCGAGCGGCTGCTCCAGGACCGGGCTGCACTGCTGACCGCG 1293
Db 1210 CGTTTCCCGGCGAGGATACGAAAGCGCTGCAAAACCTATGACCTGCTGCTGATGCCG 1269
QY 1294 ACCAGGCTACCGTTGCAACGAGCTGTGAGCGGCTCGCAAGACACCATGTCCATGATC 1353
Db 1270 ACCAGGCTACCGGCTCCACCCACCCCGCCAG---CGAATGCTCGATCAGGAGTAC 1326
QY 1354 CCAGGATGAGCGGCAATGATCTCTCAACAGCTGCGCGCTGAGACCTCAGCGGTCAACCG 1413
Db 1327 GTGGCTCGCGGCTTGGAAATGATCGGCAATACCGCGCACAGGACATCACCAGGACATCG 1386
QY 1414 GCGCTGACGCTGCGCACGGGTGCGGAGAGGCGCTGCGCGTGGCTTGGCTTCAAAGTGATA 1473
Db 1387 GCAATGCTGATC---CGTGGGCTGCTGAGCGGCTTGGCGCTGCGGTGATGCTGCTC 1443
QY 1474 GCGCGCACTTCGAGGAGTGAAGCTTACGCAACCGCGCGGCTGATGAGGCGCGCG 1531
Db 1444 GCAAAACACTACGCGGAGGACGATTTACCAAGCGCGCGGCTTGAAGCGCTCGG 1501

RESULT 3

US-09-687-594-20
Sequence 20, Application US/09687594
Patent No. 6251650

GENERAL INFORMATION:

APPLICANT: ROBERT D. FALLON
APPLICANT: MARK S. PAYNE
APPLICANT: MARK J. NELSON

TITLE OF INVENTION: NUCLEIC ACID FRAGMENTS ENCODING

TITLE OF INVENTION: STEREOSPECIFIC NITRILE HYDRATASE AND AMIDASE ENZYMES AND
SOFTWARE: MICROSOFT WORD 2.0C
TITLE OF INVENTION: RECOMBINANT ORGANISMS EXPRESSING THOSE ENZYMES USEFUL FOR
THE PRODUCTION OF CHIRAL AMIDES AND ACIDS

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESS:

ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY

STREET: 1007 MARKET STREET

CITY: WILMINGTON

STATE: DELAWARE

COUNTRY: UNITED STATES OF AMERICA

ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: MICROSOFT WINDOWS 3.1

SOFTWARE: MICROSOFT WORD 2.0C

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/687,594

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/726,136

FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA A.
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9677
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1521 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: No
US-09-687-594-20

Query Match 14.2%; Score 224.8; DB 3; Length 1521;
Best Local Similarity 48.8%; Pred. No. 7.8e-35;
Matches 731; Conservative 0; Mismatches 752; Indels 15; Gaps 4;

QY 37 CCGACAGCCGAAACAGCTCCAGGAGTACAGCGCCGCCACCACTTCGACCTCGACGAGAA 96
DB 16 CCTACCTCGACAGGTTTATAGACATCGAACCAGTTGCACATGCAACTGACGACGAA 75

QY 97 CTGGCCGCCACAGCTGTTCCGCTCGTGGCGGAGATGGTACTGCTTCGACCTGATCGAC 156
DB 76 CAGGAGCGCTCTTACTCGGAACGTGATGCAACCGAGTTTCGACGCTACGACTGGTGCAC 135

QY 157 GAACTACCCAAACCGCGCAGCGCCGCGACGCGCTACAGGACGCGCATCGGCCGCAA 216
DB 136 GAACTGGCTGATTTCTGTTCCGCCAATACGCTACGACCGCAGTTTCAGGCTATCGCCATCGG 195

QY 217 CCGACGGCGCAGAGACCGGTTTCAACGCAATTCATCCGGTTCTGCGGGGTGAGGCGCC 276
DB 196 CCATCGGCCAAGGAAACCTCTGAACCGCTTGTACTACCGAAACAGAAAGTGAATGGTGC 255

QY 277 ACGGAGGGCCACTGAGCACCTGACCGCCGCGATCAAGGACTGCTACGCTACCGCTCCCGGT 336
DB 256 CGCGAAGGCTGCTGGCGGGAACACCTGTCGCGCTCAAGATAATATCTCTTGGCAGGC 315

QY 337 ATGCCACACGAAACGGTCCCGGATGCTCCGATGTGTCGATCGCCACCGAGATGCCGTG 396
DB 316 GTCCCATGATGAACGGCGCAGCGCTTGGAAAGGCTTCTCCCGGGTTCGATGCCACG 375

QY 397 GTGGTGGAGCGCTGCTCGCGCAGCGCCACCATGTCGCGCAAGACGAACTCGAGGAC 456
DB 376 GTGGTACCCGTTGCTGATCGGGGGGACCATCTCGGCAAGGACCATCTGCGAGCAC 435

QY 457 ATGGCGAT--GGGTATCGGTGAAGCAGCGTCTACGGTCTCTGCGCTGAACCCCGAACAC 513
DB 436 TACTGCTTTCAGGAGGACGACACACCTCCGATCCAGCCCGGTGCAACAACCATCGC 495

QY 514 CCGCCACGCGACGGTGGATCTTCCAGCGGCTCGGGGCTGCGCTGCTGCGCGCATG 573
DB 496 CACGGTTATGCTCTGCGGGTTCCTCATCAGGACGCGGGCATTTGGTTCGCTCGCGTGA 555

QY 574 GTGACTTCGCTCGGCGCTCGATGAGGAGGAGCAGCATCCGGATCCCGCCCGCATGTGTC 633
DB 556 GTGGACATCGCTGGCGCGCATCAAGCGGCTCCATTCGATCCCGTCCGCTTCTGCG 615

QY 634 GGAAGTGGTGGATGAGGCGGACCGCGCTGGTGGCTTACGGCTGACATACATG 693
DB 616 GTTACTACGGATGAAGCCACCCACCGGCTGGTGCCTACACCGGGCTCATGGCGATT 675

QY 694 GACCACACCTTGGACACCATCGGGCCCATCAGCGGGGGTTCGAGTCAACCGCCGGGTC 753
DB 676 GAAGCCACGATCGATCATCTCGGCCCATCACCGGTAACTGCGGCAACCGCGTGTATG 735

QY 754 CTCGAGGTGTTGCCGGGCGCGACTGGCGGACCTTCAGTGGGTGCTTAACCTTCGGAG 813
DB 736 CTCGAGGCAATGGCCGGTGCAGCGGACTCGACCCCGCGCGCGCGGCGGCGGCGGCGG 789

QY 814 CCGGAGAACTACGGCTCCGGCTCGCGGAGGAGTATCCGGTCTGAGATTTCGGGTGTC 873
DB 790 GTGATGACTATTTCAGTTACTTGGAAAAGGCTGAGCGGACTCAGATCGGGTGTG 849

QY 874 GAGGAGTCACTGAGCGGAAACGGTGGAGCGCGGAGTATGCCGCTTCAACAGGGA 933
DB 850 CAAGAGGGAATTCGGCTTGTAAACAGGACCTTCGCTGGCGGACAAAGTGCAGCGCC 909

QY 934 CTGGCGCGCTCGAGAGCGCGCTGCGACCATCAGCGGGGTCTCGGTGCGCTTGTGAGC 993
DB 910 ATGCGCGACTCGAGGGTTGGCGCTCATGTGAGCGGCTTCCATTCGAGGACAAAC 969

QY 994 GCGGCTTGGCTATCCAGAGCGCGCTGATGGCTTTCAACGCGCGCTATGCGGACTCC 1053
DB 970 CTGGCAGGGTGTGTGGCACCCCATCGGTTCGGAAGGCTTGACATGATGATGCA 1029

QY 1054 GCGGTGTGGCTACTTCCACAGGGGCGCTGAGAGTCAAGACCGCTCAGACGGCG 1113
DB 1030 GGCAACGGCGCAGGCTTTAACTGGAAGGACTTTACGATGTGCGCTCTGCAATGTTGTC 1089

QY 1114 GCCAGAGTGCACACCCCAACAGGATCTGGCGATCTCTGTCGCGCTGATGTGTTGATC 1173
DB 1090 GCCAGCTGGCGGACGACGAGCAGACCAATATTCGCGCTCAAGCTCTGCAATGTTGTC 1149

QY 1174 GCGGAGCACTGCGCGCAGGATACCTCGGCATCCATAGCGGAGGCGGAGACCTGCGG 1233
DB 1150 GGGCAATACGGCTGTGCGGCTACACGAGCGCTACTACGCGCAAGGCGCAGAACCTTGA 1209

QY 1234 CTGGAGTCTGGCAAGCAGATTCGACGCGCTCTCCAGGACGGGCTGCATCTGTGACCCCG 1293
DB 1210 CGTTTTCGCGGAGGATACGACAAAGCGCTGCAAACTATGACCTGCTGGTGTGTCG 1269

QY 1294 ACACGGCTACCTGTCGCAAGAGCTGTGAGCGCTGCGGAGGAGAGGCTTGGCTTCAAGT 1353
DB 1270 ACCAGCGCTACGCGCCCAACCCCAACCCCGCAG---CGAACTGCTCGATCACGGAGTAC 1326

QY 1354 CCAGGATGACGGGCAATGCTCAACAGCTGCGCTGAGCTCACCGCTCACCCCG 1413
DB 1327 GTGGCTCGCGGTGGAAATGATCGGCAATACCGCGCACAGACATCACCGGCTCCG 1386

QY 1414 GCGTGAAGTGGCCACGGGTGCGGCGGAGAGAGGCTTGGCTTGGCTTCAAGTATGATA 1473
DB 1387 GCGATGTGATTC---CGTGTGGCTGCTGGACGGCTGCGCTGCGGTGATGCTGTC 1443

QY 1474 GCGCGCACCTTCGAGGAGTGCAGCTCTACCGCACCGCGCGCTGATCAGAGCGCGCG 1531
DB 1444 GCAAAACACTACCGCGGAGGAGGAGGATTTACCAAGCGCGCGGCTTTGAGGCTCGG 1501

RESULT 4
US-08-539-666-1
Sequence 1, Application US/08539666
Patent No. 5766918
GENERAL INFORMATION:
APPLICANT: Petre, Dominique
APPLICANT: Cerbelaud, Edith
APPLICANT: Mayaux, Jean-Francois
APPLICANT: Yeh, Patrice
TITLE OF INVENTION: No. 5766918el Polypeptides, The DNA Sequences
TITLE OF INVENTION: Allowing their Expression, Method of Preparation, and
TITLE OF INVENTION: Utilization
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/539,666
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/097,009
FILING DATE: 27-JUL-1993
APPLICATION NUMBER: US 07/612,673
FILING DATE: 14-NOV-1990
APPLICATION NUMBER: FR 8916332
FILING DATE: 11-DEC-1989
ATTORNEY/AGENT INFORMATION:
NAME: Potter, Jane E.R.
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 03715.0010-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1879 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-539-666-1

Query Match 11.5%; Score 180.8; DB 1; Length 1879;
Best Local Similarity 48.1%; Pred. No. 2.2e-26;
Matches 704; Conservative 0; Mismatches 742; Indels 18; Gaps 6;

Qy 66 GCCTGCTGCTTACGCGCTGACATGGAACACACCTTGGACACATCTCGGCGCCAT 722
Db |||||
Qy 883 GCTGCTCCGTAATACCGGTGCTATCCCATCGAGCAACATGACCATCTCGGCGCAT 942
Db |||||
Qy 723 CACACGGGGGTGAGCTCAACGCGGGTCTCGAGGTGTTGGCGGGGCGGAGTGGG 782
Db |||||
Qy 943 CACACGGCGGTCCAGCATGATGCTCTCGGTCTATCGCGGCGCGGCGCGGTA 1002
Db |||||
Qy 783 CGACCTCAGTGGGTGCGTAACCTTCGCGAGCGCGAGAACTACGCGCTCCGCGCTCGGCGA 842
Db |||||
Qy 1003 CGACCCAC---GCCAAGCGGACAGTGTGAGAGAGGTGACTATCTGCTCCACCTCGACTC 1059
Db |||||
Qy 843 GGGAGTATCCGGTCTGAGATTTCGGGTGCTGAGAGAGTCACTGAGAGCGGAACGTTGGAC 902
Db |||||
Qy 1060 CGATGTGAGCGGCTCGGAATCGGAATCGTTCGAGAGGGATTTCGGGCGACCGCGTCTCACA 1119
Db |||||
Qy 903 GCGGAGCTGATCGCGCGGTTCACAGGGGACTCGCGCGCTCGAGAGCGCGGTGCGAC 962
Db |||||
Qy 1120 GCGGAGGTGAGAGCGAGTTCGCGGAGCGGACAGTCTGACCGAATCGGTGCGAC 1179
Db |||||
Qy 963 CATGAGCGGTCTCGGTGCGGTGTTGGAACGCGCGCTGCTATCCAGAGCG---GGGT 1019
Db |||||
Qy 1180 GGTAGAGGAAGTAAACATCCCGTGGCATCTGCATGCTTCCACATCTGGAACGTGATCGC 1239
Db |||||
Qy 1020 GATGCTTTCAACGCGCGGTATGCGGAGTCTCGCGCGTGTGGCTACTTCCACAGCG 1079
Db |||||
Qy 1240 CACGAGCGGTGCTTACAGATGTTGGAACGGAACGATACGGCATGAACGCGAAGG 1299
Db |||||
Qy 1080 GCGGTGAGCGTCAAGCAGCGCGCTCAACGAGCGCGCCAGAGTCCGACCCACCAAGGA 1139
Db |||||
Qy 1300 TTGTGACGATCGGAACATGATGGGACACTTTGCTTCTCGAGCATTCAGCACGCGGACGC 1359
Db |||||
Qy 1140 TGTGCGGATCTGTCCCGCTGATGCTGCTGATCGGAGGACCTGCGGAGCGATACCT 1199
Db |||||
Qy 1360 TCTGTCCGAACCGTCAAACTGGTGGCTGACCGGCGCCACCCAGGCGATCACCCCTCGG 1419
Db |||||
Qy 1200 CCGCATCCACTACGGAAGGCGGAGAACTCGCGGTGAGCTCGGCAAGAGATCGAGCGC 1259
Db |||||
Qy 1420 CCGCGGAGTACGCAAGAACCGGAACTCGTACCGCTTGCCTGCGCGCGCGCTACGACAC 1479
Db |||||
Qy 1260 GGTCTCCAGGACCGGCTGACTGCTGACCGGACGCGCTACCTGTCGCAAGAGCT 1319
Db |||||
Qy 1480 TGCCTTGAGCAATTCGACGCTCTGGTGTATCCCAACGCTGCGCTACGTCGCAAT 1539
Db |||||
Qy 1320 GTTGAGCGTCCGGAAGACACCATGCTCCATGATCCACGATGACGGGCAATGCGCT 1379
Db |||||
Qy 1540 GCGGCGAAGACGTAGATCGTGCACTTATCAC---CAAGGCTCTCGGATGATCGC 1596
Db |||||
Qy 1380 CAAACGTCGCGCTGAGCTCAACGCTGACCGGCGGTGAGCGGTCGCGGCGG 1439
Db |||||
Qy 1597 CAAACGCGCACCATTCGACGTCGAGCGGACATCCGCTCCCTGCTCGGCGCGGCTG--- 1654
Db |||||
Qy 1440 CGAGAAGCGCTGCGCGTGGCTCCAAAGTATAGGCGCGCACTTCGAGGAGTGCACGCT 1499
Db |||||
Qy 1655 -GTGAACCGGCTTCGCTCGGATGATGATCACCGGAGACACTTCGAGTGCACGCT 1713
Db |||||
Qy 1500 CTACCGCGCGCGCGCTGATCGA 1523
Db |||||
Qy 1714 CTTCTGTCGAGCGGCTGATCGA 1737
Db |||||

RESULT 5

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R. M.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS

```
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      11.0%; Score 172.8; DB 3; Length 4403765;
Best Local Similarity 48.9%; Pred. No. 1.5e-24;
Matches 623; Conservative 2; Mismatches 619; Indels 30; Gaps 5;

QY 275 CCACGAGGGGCGCACTGAGCGACCTGACCGCGCATCAAGGACTGCAATGCCATCGCGC 334
Db 3365561 CTTTCCCGTCCGCGCTGGCCGGGTGCGCGTCAAGGACGTCTTACACACGCG 3365502

QY 335 GTATGCCACACGAAAGCGTCCCGATGCTCCCGATGTTGATCGCCACCGAGATGCG 394
Db 3365501 ACATGCCACACCTCGCGGTCAAAATCTGGAGGATGCGATCTCCCTACGACGCA 3365442

QY 395 TGGTGTGAGCGGTGCTCGCGGAGCGGCCACCATCGTGGCAAGACGAACTCGAGG 454
Db 3365441 CGCTGACCGCGGTGCGCGCGGGGATCCGATCTCTGGCAAGACCAATGAGCG 3365382

QY 455 ACATGCGCATGGG---TATCGGTGAAGCAGCGTCTACGGTCTCGCTGAACCGAACA 511
Db 3365381 AGTTGCGATGGGCTCGTCAAGGAGAACTCCGTTACGGTCCCAACCGCAACCGTGA 3365322

QY 512 ACCCGCCACGCGCGGTGATCTCCAGCGGTCCGGCGCTCGCTCGCTCGCGCA 571
Db 3365321 ATCTGACCGGTATCCCGCGGTTCGGTGGCGGAGCGCGCGCGTTCGCGCGTTCC 3365262

QY 572 TGGTGCATTCGCTCGCGCTGATGAGGAGGAGGAGGATCGGATCCCGCGCGCATGGT 631
Db 3365261 AGCGCGCTGCGCATGATGAGGAGGAGGAGGATCGGATCCCGCGCGCATGGT 3365202

QY 632 GCGACTGTCGCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 691
Db 3365201 CCGGACCGTGGCGGTCAAAACCACTACGGGACCGGTGCGCGTATGGGTGGGCT 3365142

QY 692 TGGACACACCTTGACACCATCGGCGCCCATCACAGGAGGAGGAGGAGGAGGAGGAGG 751
Db 3365141 GCGGTCTCTGCTGATCAGGCGGCGCGCTGTGCGCGCACCGTCTTGACACCGCGTGT 3365082

QY 752 TCTCGAGGTGTTGGCGGGGCGGACTGCGCGGACCTCAGTGGGTGGTAACTTCCGG 811
Db 3365081 TGATCAGGTGATCGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3365022

QY 812 AGCGGAGAACTACGCTCGCGCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 871
Db 3365021 CCGAGTGTGGGCGCGCTAGGCGGCGGCGGCTGGGATCTGCTGGGTGGGAGGAGGAGGAGG 3364962

QY 872 TGAGAGGTACTTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 925
Db 3364961 GCGTGTTCGACAGGTGACGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3364902

QY 926 ACCAGGAGTTCGCGGCGCTCGAGAGCGCGGTCGCGGAGGAGGAGGAGGAGGAGGAGG 985
Db 3364901 AGGTCGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3364842

QY 986 TGTGAGCGGCGCTGGCTTATCAGAGCGGCGGTGATGCTTCAACG---CGCGCGGTA 1042
Db 3364841 ACTTCGACCATGCGCTGGCGGCTTATACCTGATTCGCTCGGAGGTGTCGAGCAATC 3364782

QY 1043 TGGCGAGCTCCGCGGTGGGCTACTTCCAAAGGCGGCGGTGAGCACTCAGCACCGCGC 1102
Db 3364781 TGGCGAGCTCCGCGGTGGGCTACTTCCAAAGGCGGCGGTGAGCACTCAGCACCGCGC 3364722

QY 1103 TCACGAGCGCGCGCCAGAGTTCGACACACCAAGGATCTGGCGATCTCTGTCCCGCTGA 1162
Db 3364721 CCAGGAGGTGATGCGGATGACCCGCGCGCGCGGTTTCGGGCGCGAGGTCAAGCGCGCA 3364662

QY 1163 TGGTGTGATCGCGGAGCAGCTGCGGAGCAATACCTCGGCATCCACTACGGAAGGCGC 1222
Db 3364661 TCATGATCGGACCTACGCGTGTTCGGCGGCTACTACGACGCTATTATCAACACGAGCGC 3364602

QY 1223 AGAACCTTGGGCTTGGAGCTCGGCAAGCAGATCGACCGCGCTCTCTCCAGGACCGGCTGCA 1282
Db 3364601 AGAAGTGGCGACGCTGATCGCGCGGACCTCGAGCGCGGATCGTCCGCTCGAGGTGC 3364542

QY 1283 TGGTGAACCGGACGACGCTTACGTTGCAACGAGCTGTGAGCGGTGCGGAGACCA 1342
Db 3364541 TGGTGTGGCGGACGACGCTTACGTTGCAACGAGCTGTGAGCGGTGCGGAGGTGCGGAGGTGC 3364482

QY 1343 TGTCCATGATCCACGAGTACGCGGCAATGCGATCTCAACGCTGCCCGCTGGACCTCA 1402
Db 3364481 TGGCGATGATCTGTTGCA-----CCTGTGCACTGCGCTGAACTTGG 3364437

QY 1403 CCGGTACCGCGCGCTGACGTTGCCCAACGCGTG---CGGCGGAGAGGCGCTGCGCGTTG 1459
Db 3364436 CCGGCGCTGCGGCTGCTGTGCGCTCGGCGCTCTCCCGGACGAGCGGTGCGCGTTG 3364377

QY 1460 GCCTCAAGTGTAGGCGCGCACTTCGAGAGTTCGAGCTCTACCGCAACCGCGCGGTGA 1519
Db 3364376 GCCTACAGATCAAGCGCGGCAATGCGCGAGCGGCTCTACCGGTGGGCGGCGCTT 3364317

QY 1520 TCAGGCGCGCGC 1533
Db 3364316 ATCAGGCGCGCGC 3364303

RESULT 6
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      10.8%; Score 170.4; DB 3; Length 4411529;
Best Local Similarity 48.9%; Pred. No. 4.4e-24;
Matches 623; Conservative 0; Mismatches 621; Indels 30; Gaps 5;

QY 275 CCACGAGGGGCGCACTGAGCGACCTGACCGCGCATCAAGGACTGCAATGCCATCGCGC 334
Db 3371229 CTTTCCCGTCCGCGCTGGCCGGGTGCGCGTCAAGGAGGAGGAGGAGGAGGAGGAGG 3371170

QY 335 GTATGCCACACGAAAGCGTCCCGATGCTCCCGATGTTGATCGCCACCGAGGATGCG 394
Db 3371169 ACATGCCACACCTCGCGGTCAAAATCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3371110

QY 395 TGGTGTGAGCGGTGCTCGCGGAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 454
Db 3371109 CGCTGACCGCGGTTGCGCGCGGAGTCCCGATCTCTGGCAAGACCAATGAGCG 3371050
```

QY 455 ACATGCGATGGG---TATCGGTGAAGGAGAGCTCTACGGTCTCGGCTGAACCCGGAACA 511
Db 3371049 AGTTCGGATGGCTCGTACAGAGAACTCCGCTTACGGTCCACCCGCAACCCGTGA 3370990

QY 512 ACCCGCCACAGCAGCGGTGATCTTCAGCGGCTCCGGCTCGGCTCGCTCGCGGA 571
Db 3370989 ATCTCAGCGGTACCGCGGCTTCGCGTGGCGGAGCGCGGCGCTGGCGCGGTTC 3370930

QY 572 TGTGACATTCGCGCTCGGCGTTCGATGAGGAGGAGCAGATCCGGATCCCGGCGCGATGT 631
Db 3370929 AGCGCGCTTCGCGATCGATCCGACACCGGCGGTTCGATCCCGCAGCGCGCGCTGA 3370870

QY 632 GCGGACTGTGCGCATGAAGCGGACCCAGCGGCTGTGCGCTTACGGCTTCGACATACA 691
Db 3370869 CCGGACCGTGGCGGTCAACACCCACTACGCGCGGTGTGCGCTATGGCTGGTGGCT 3370810

QY 692 TGACACACACTTGGACCAATCGGCGCCATCACAGGGGGTTCGAGCTCAACGCGCCGG 751
Db 3370809 GCGGCTCTCGTGGATCAGGCGCGCGCTGTGCGCGCACCGCTCTTGGACACGGCGTGT 3370750

QY 752 TCTCAGAGTGTGGCGCGGCGGCTGAGCGGCGGCGCTGCGGATCTGCGTGGCGGTTCG 3370630
Db 3370749 TGCATCAGGTGATCGCGCGGCGGAGCGCGGCGGCTTCCAGCTGCTGCGAGCGGAGTGC 3370690

QY 812 AGCGGAGAACTACGCTCGCGCTCGGCGAGGAGTATCCGCTCTGAGATTCGCGGCTCG 871
Db 3370689 CCGAGTGTGGCGCGCTAGGCGCGCGCGCTGCGGATCTGCGTGGCGGTTCG 3370630

QY 872 TCAGGAGTCACTGAGCGGAGCGGTGCG-----ACGCGGAGGTGATCGCGCGGTTC 925
Db 3370629 GCGTGGTTCAGAGTCACTGCGCGGCGGAGGCTTACCAGCGGCGGTGCTGGCTCTCTCG 3370570

QY 926 ACCAGGAGTGGCGGCGCTCGAGCGCGGCTGCGACCATCGAGCGGTCTCGGTGCGT 985
Db 3370569 AGCTGCGGTGAGAGCTAACCGCGCTGCGCGCTGAGTCAAGGAGTCACTGCGCGCT 3370510

QY 986 TGTGAGCGGCGCTCGCTTATCCAGAGCGCGGTGATGGCTTTCAACG---CGCGGCTA 1042
Db 3370509 ACTTCGACCATGCTTGGCGCGCTATTACCTGATCTGCGCTCGGAGGTTCGAGCAAT 3370450

QY 1043 TGGGAGTCCGCGGTGTGGGTACTTCCAGAGGCGGCGGTGAGCTCAGCAGCGCG 1102
Db 3370449 TGGCGGCTTCGACGCGGATCGCTACGCGGTGCGGCTGCGGAGCAGCAGCAGCGG 3370390

QY 1103 TCAGACGCGCGGCGAGTTCGACACACCCAGGATCTGGGATCTGCGGTCTCGGCTGA 1162
Db 3370389 CCGAGGAGGTGATGGCGATGACCGGCGCGCGGTTTCGGGCGCGAGTCAAGCGCGCA 3370330

QY 1163 TGCTGTGATCGCGAGACCTCGCGGAGCAATACCTCGGCTATCTAGCGGAGGCGG 1222
Db 3370329 TCATGATCGGACCTACGCGTTCGCGCGGCTACTACGAGGCTATTACACAGCGCG 3370270

QY 1223 AGAACCTCGCGCTGAGCTCGGCAAGCAGATCGACGCGGCTCTCCAGGACCGGCTGAC 1282
Db 3370269 AGAAGTTCGCGACGCTGATCGCGCGGACCTCGACGCGGCTATCGTCTGCTCAGTGC 3370210

QY 1283 TGCTGAACCGGACACCGCTTACCGTTGCGCAACGAGTGTGTGAGCGGTGCGGCAAGACCA 1342
Db 3370209 TGGTGTGCGGCGACGACCGGCGGCTTCCGATGAGAGGTGAGAGGTGAGAGTCCG 3370150

QY 1343 TGTCATGATCCGAGGATGAGGCGCATCGATCTCTCAACAGTGGCGGCTGACCTCA 1402
Db 3370149 TGGGATGTACTTGTTCGA-----CCTGTGACGCTGCGGCTGAATTGG 3370105

QY 1403 CCGTCAACCGGCGCTGACGCTGCGGCGGCGG---TGCGGCGGAGAGGCGCTCGCGGTTG 1459
Db 3370104 CCGGCGACTGCGGCTGCTGTGCGGCTGCGGCTCTCCCGGAGCAGGCTTCCGCGTTG 3370045

QY 1460 GCCTCGAGTGAAGGCGGCACTTCAGAGAGTTCGAGCTCTACCGCACCGCGGCGCGTGA 1519
Db 3370044 GCCTACAGATCATGGCGCGGCAATTGGCGGAGCAGCAGCGGCTCTACCGGCTGCGGCGGCTT 3369985

QY 1520 TCAGGCGCGCGCG 1533
Db 3369984 ATGAGCGCGCGCG 3369971

RESULT 7
US-07-612-673-1
; Sequence 1, Application US/07612673
; Patent No. 5260208
; GENERAL INFORMATION:
; APPLICANT: Petre, Dominique
; APPLICANT: Cerbelaud, Edith
; APPLICANT: Mayaux, Jean-Francois
; APPLICANT: Yeh, Patrice
; TITLE OF INVENTION: NOVEL POLYPEPTIDES, THE DNA SEQUENCES
; TITLE OF INVENTION: ALLOWING THEIR EXPRESSION, METHOD OF PREPARATION, AND
; TITLE OF INVENTION: UTILIZATION
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/612,673
; FILING DATE: 19901114
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E.R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 03715.0010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1878 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-07-612-673-1

Query Match 10.7%; Score 168.8; DB 1; Length 1878;
Best Local Similarity 48.0%; Pred. No. 4.4e-24;
Matches 703; Conservative 0; Mismatches 742; Indels 19; Gaps 7;

QY 66 CGCGCGCCACCATTCGACCTCGAGGAGTCCGCGCCAGCTCGTTCGGTCTGTCG 125
Db 286 CGAAGGATTCAGGATCACTCTCGAACAACGCCGCTCGAGTGGCGGCACTGAT 345

QY 126 GGAGATGTGATCGCTTCGACCTGATCGAGAACTACCGCAACCGCGCAGCCGCGAC 185
Db 346 CGACGAGCACTGGGCTCTAGCAGTCTCGACAGTTGTACCGCAGAGGCGACCCC 405

QY 186 GCGGTACAGGACCGGACATCGCGCGGCAACCGACCGGCGGCGGAGACCGGTTCAAGC 245
Db 406 GCGGACCACTGACCGGAGCAG---CGTCCCAAGTTCGAGCGGAAATCTTTGAGCGC 462

QY 246 ATTCATCCGTTCTGCGGCTGGAGGCGCCACAGAGGCGGCACTGAGCGACCTGACCGC 305
Db 463 TTGGTATGTGACCACCACTCCCGCGACGCTCGGAGCGGCTCTGACCGCGCGAGCGT 522

QY 306 CCGCATCAAGAGTCACTGCGGCTGCGGCTGATCGGCTATGCGGCGGCGGCGGCGGATGCT 365
Db 523 GCGGATCAAGGACACGTCGAGCGGCTGCGGAGTTCGCGATGATGACCGGATCTCGGACGCT 582

QY 630 CATGACACACCTTGGACCACTCCGGCCCATCACAGGGGGTTCAGCTCAACGCCG 749
Db 621 CTACCTTCCAGCTTCGACCAAGGGGGCCCTGGCGCGCAACCCGAGAGACTCGCGCT 680
QY 750 GGTCTCTAGAGTGTGGCGGGCGGCGAGTGGCGGACCCCTCAGTGGGTGCGTAACTTCC 809
Db 681 GATGCTGGGGTGTATGGCGGATTCGATCCGAGGACTCGACAGCTT-----CGACA 734
QY 810 GGAGCGGAGACTACGCTCGCGCTCGGCGAGGGAGTATCCGCTTGAGATTGGCGT 869
Db 735 GCGGTGAGACACTACCTTGGCGGCGCTCGAAGAGCGCTGAGCGGCTCGGCATCGCCT 794
QY 870 CGTCAGAGTCACTGGAGCGGAACGGTGGCAGCGCGGACGTATCGCGGTTCAACCA 929
Db 795 GCGCGGGAATATTTCGCGCGGCTCGACAGCGCATCGCGAGCGGTCTGGCC-- 852
QY 930 GGGACTGGCGGCTCGAGAGCGCGGTGGACCATCGAGCGGTCTCGGTGCGGTGTG 989
Db 853 -GTAGTCAGAGCTGAAGACGCTCGGCGCCACGGTGAAGACATTTCCCTGCGGAACAT 911
QY 990 GACGCGGCTGGGCTATCCAGAGCGGGGTGATGGCTTTCAACGCGCGCTATGGCGGA 1049
Db 912 GCAGACGCCATCCAGCCTACTAGTGTATCGCGCGCGCGAGGCTCTCAACCTGTC 971
QY 1050 CTCGCCCGTGTGGCTATTCCACAGGGGCGGTGACGTCAAGACCGCGCTCAACAC 1109
Db 972 GCGCTTCAGCGCGTGTGCTATGGCTATCGTTGCGACCGCGCGGAGACATTTCCCTGCGGAACAT 1031
QY 1110 GCGCGCCAGAGTGCACACCAAGGATCTGCGGATCTGCGGATCTGCGGCTGCTGGT 1169
Db 1032 GTACAGCGCTCGCGCGGAGGCTTCGGCAGCGAGTGAAGAACGCTGGAAGACCT 1091
QY 1170 GATCGCGAGACCTGCGCGAGCAATACCTCGGCATCCACTACGCGAGGCGCGAGAACCT 1229
Db 1092 CACTTACGACTCTCG--GCGCGCTACTAGATGCTTATACCTGCGAGGCTCAGAAGAT 1148
QY 1230 GGGGTGGAGTGGCAAGCAGTACGCGCGTCTCCAGACCGGGTGCACCTGCTGAC 1289
Db 1149 TCGCGGTGATCAAGACGACTTCTGTCAGCGCTTTCCGAGTGGACGTGCTCTCGG 1208
QY 1290 CCGGACACGCTACCTGTGCGCAAGCTGTGAGCGGTGCGCAAGACACCATGTCCAT 1349
Db 1209 CCGGACACGCGCAACCGCGCTGGAAGATCGCGGAGAGAACGACGACCGGTTTCCCA 1268
QY 1350 GATCCACGAGTACGGGCAATGCGATCTCAACAGTGGCGGCTGAGCTCACCGGTCA 1409
Db 1269 GTACCTG-----GAAGACATCTACACCATCACCGCAACCTCGCGGCT 1313
QY 1410 CCGCGGCTGACGCTGCGGCGGAGGAGGCGGCTGCGGCTGCGGCTGCGGCTTCCAGT 1469
Db 1314 GCGGCGGTGTCCATGCGCGCGGCTTCTG--ACGGCTGCGGTGGGTGCTGAGTT 1370
QY 1470 GATAGCGGCACCTTCAGGA 1490
Db 1371 GCTCGCGGCTACTTCCAGGA 1391

RESULT 10

US-09-252-991A-10084/c
; Sequence 10084, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10084
; LENGTH: 1818
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10084

Query Match 9.6%; Score 151; DB 4; Length 1818;
Best Local Similarity 49.0%; Pred. No. 1.1e-20;

Matches 598; Conservative 0; Mismatches 590; Indels 33; Gaps 6;

QY 273 CGCCAGGAGGCGCCACTGAGCGACCTGACCCCGCGCATCAAGGACTGTCATCGCATCGC 332
Db 1612 CGCGAGAAACGCGCCCTCTCGCGCGCGCATGCCACAGGAGACCTGTTCTGCACCCA 1553
QY 333 CGGTATGCCACACACGACGCGGTCCCGATGCTCCGAGTGTATGCGACCGAGATGC 392
Db 1552 GGGCGTACGACACGAGTTCGAGATGTCGACCAACTTCGTCCTCGCCCTCAGAGC 1493
QY 393 CGTGTGTGTGAGCGGCTGTCGCGCGAGCGCCACCATCGTCGCGCAAGCAACTCGA 452
Db 1492 CACCGTGTGAGAGAGCTCACCGCGCGCGCGCTTACCTTCGGCAGCTGAACATGA 1433
QY 453 GGCATGCGGATGGG---TATCGGTGAAGGAGAGGTCTACGGTCTCTGCTGAAACCCGAA 509
Db 1432 TGAATTCGCCATGGGCTCGTGAACCCAGTCCAGCCACTACGCGCGGTGAAGAACCCCTG 1373
QY 510 CAAACCGCGCCACGCGCACGCGGTGATCTTCAGCGGCTCCGCGCTGCGTCTGCGCG 569
Db 1372 GAGCTTCGACCGGTGCGCGGCTCTCCGCGGTTCGCGCGGCGAGTTCGCGCGG 1313
QY 570 CATGTGTCGATCTTCCCTGGCGTTCGATGAGCGAGCAGCATCGGATCCCGCGCGCATG 629
Db 1312 CTGTGTCGCGCGCGCACCGACGATACCGCGCGCTCGATCCGCAACCGCGCGCGCT 1253
QY 630 GTGCGGACTGTGCGGATGAAGCGACCGACCGCGCTGCGGTCTTACGGCTGACATA 689
Db 1252 GACCAACTTCACGGGATCAAGCAACCTACGCGCGGTTCGCGTGGGATGATCGG 1193
QY 690 CATGACCAACACTTGGACCACTCGGCGCATCGGCGCATACAGGGGGGTGAGCTCAAGCCG 749
Db 1192 CTACGCTTCAGGCTCGACGCGCGCGCGCTGCGCGCGCGCGCGCGAGGACTGCGCGCT 1133
QY 750 GGTCTCGAGGTTGGCGCGCGCGCGGCTGCGCGCGCGCTCAGTGGGTGCGTAACTTCC 809
Db 1132 GATGCTGGGGGTGATGGCGGATTCGATCCGAGGACTCGACAGCGT-----CGACA 1079
QY 810 GAGCGCGGAGAACTACGCGCTCGCGCTCGCGAGGAGTATCCGCTCTGAGATTGCGGT 869
Db 1078 GCGGTGAGGAGTACTGCGCGCTCGAGAAAGCGCTGAGCGGCTCGCATCGGCT 1019
QY 870 GTCGAGGAGTCACTGAGCGGAAAGCGGTGCGAGCGCGGAGTATCGCGCGCTCAACCA 929
Db 1018 GCGCGGGAATATTTCGCGCGCGGCTCGACAGCGCATCGCGCGCGGTCTGCGCC-- 961
QY 930 GGGACTTGGCGGCTCGAGAGCGCGGTGCGACCATCGAGCGGCTCTCGGTGCGGTGTG 989
Db 960 -GTAGTCGAGGAGTGAAGACGCTCGCGCGCACGCTGAGGACATTTCCCTGCCGAACAT 902
QY 990 GACGCGGCTGCGCTATCAGAGCGGCTGATGGCTTCAACGCGCGGCTATGCGGA 1049
Db 901 GCAGACGCCATCCAGGCTTACTGATGTCGCGCGCGGCGGTCTCTCAAACTGTC 842
QY 1050 CTCGCGCGGTGGGCTACTTCCAAAGGGCGCGTGAAGCTCAGCACCGCGCTCAGCAC 1109
Db 841 GCGCTTCGAGCGGTGCGCTATGGCTATGCTTGGAGCGCGCGCAGAACTGGAAGACCT 782
QY 1110 GCGCGCGGAGTTCGCAACCAAGGATCTGGGATCTGCTGCTGCTGCTGCTGCTGCT 1169
Db 781 GTACAGCGCTCGCGCGGAGGCTTCGCGAGCGAAAGTGAAGAACCGCATCATGCTCGG 722
QY 1170 GATCGCGGAGCACTCGCGGAGCAATACCTCGGCATCCACTACGGAAGCGCGCAACT 1229

Db 721 CACTACCGACTCTCG---CGCGGCTACTAGATGCGCTATTACTCGAGGCTCAGAGAT 665
Qy 1230 GCGCTGAGCTCGGCAAGCAGATCGACGCGTCTCTCCAGAACCGGCTGCACTGCTGAC 1289
Db 664 TCGCGGCTGATCAAGAACTCTGTCAGCGCTTTGCGGAAGTGAAGTATCTCTCGG 605
Qy 1290 CCGACCAACGCTACGTTGCGCAAGAGCTTTGAGGGTCTGCGGAGACACCATGTCAT 1349
Db 604 CCGACCAACGCGCAACCCGCGCTGGAAGATCGGCGAGAAACGACGACCCGGTTTCCCA 545
Qy 1350 GATCCCCAGGATCGGCGCAATGCGATCTCTAACACGTCGCCCTGACCTCACCGGTCA 1409
Db 544 GTACCTG-----GAAGACATCTACACCATCACCGCAACCTCGCCGCGCT 500
Qy 1410 CCGCGGCTACGCTGCGCAAGGCTGCGGCGAGAGGGGCTCGCGCTTCCCAAGT 1469
Db 499 GCGGGGCTGTCCATGCGCGCGCTTCTGTCG---ACGGCTCTCGGCTGCGTCCAGTT 443
Qy 1470 GATAGCGCGCACTTCAGGA 1490
Db 442 GCTCGCGCTACTTCCAGGA 422

RESULT 11

US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 7.8%; Score 123.4; DB 3; Length 4403765;
Best Local Similarity 53.2%; Pred. No. 4.6e-15;
Matches 309; Conservative 0; Mismatches 266; Indels 6; Gaps 2;
Qy 262 CCGGTGAGGGCGCCACGAGGGGCCACTGAGCAGCTGACCGCGGATCAAGGACTGC 321
Db 2641838 CCGGGGCGCGCGGACACGCGCGCTGTGGGCACTCGGATCGGGTCAAGGACGAC 2641897
Qy 322 ATCGCATCGCGGTATGCCACCAAGAACGGGTCCCGGATGCTCCGACTGTGATCGCC 381
Db 2641898 GTGACGTTGTGGAGTGCACACCGCTTCGGCACCCAGG---CTATGTCCGCGCTGCT 2641954
Qy 382 ACCGAGGATGCGTGGTGGTGGAGCGGCTGTCTCGCGGAGGCGGCACCATGTCGGCAAG 441
Db 2641955 ACCGACGACTGTAGTGTCTCGGCGCTCAAGGCGCGGAGCGGTGATCTCGGCAAG 2642014
Qy 442 AGAACCTCGAGGACATGG---CGATGGGTATCGGTGAAGGAGCGGTCTACGGTCTTGG 498
Db 2642015 ACGAATCTTGTGAATTGGGCGAGTGGCGCTTACACGAGACCGGGTTCGGACACACC 2642074
Qy 499 CTGAACCCGCAACACCCGCCACGCGGTGATCTTCAGCGGCTCGGCGCTGCC 558
Db 2642075 CGCAACCCCTGTCTGGCGCGGACACGCGGGTGGATCTCTGGGCGGTAGCGCGCGG 2642134

Qy 559 GTCGTGCGGCATGGTTCGACTTCCCTCGGCTCGATGAGGAGGCGAGCATCGGATC 618
Db 2642135 GTGCGCGCGCTGGTTTACCGCGCTATCGGCTCCGACGGCGCGGAGCATCCGATC 2642194
Qy 619 CCGGCGCATGGTTCGGACTGTCGGCATGAAGCGGACCCACGCGCTGTCGCGCTTTTAC 678
Db 2642195 CCGGAGCATGGACACACACTAGTGGGCATCAAGCACACGCGTTCGATCTCCACCTGG 2642254
Qy 679 GGCCTGATATATGAGACACACTTGGACCATCGGCGCCCATCACCAGGGGGTTCAG 738
Db 2642255 CCGCTCGCGGAGGCGTTCAACGCGGTACAGGTCAACGCGCTACTGGCCCGCACTGTGAG 2642314
Qy 739 CTCACGCGCGGCTCTCGAGGTGTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 798
Db 2642315 GATCGGCGCTGTGCTCTGACGCGCGCTCCGCAACGTCGAGGGGCGGCGGCGG 2642374
Qy 799 CGTAACCTTCGCGAGCGAGAACTACGCGCTCCGCGCTCG 839
Db 2642375 CCCCCGCTGACGGTGTCCGATTTCGTGCGCATCGCCCTGG 2642415

RESULT 12

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 7.8%; Score 123.4; DB 3; Length 4411529;
Best Local Similarity 53.2%; Pred. No. 4.6e-15;
Matches 309; Conservative 0; Mismatches 266; Indels 6; Gaps 2;
Qy 262 CCGGTGAGGGCGCCACGAGGGGCCACTGAGCAGCTGACCGCGGATCAAGGACTGC 321
Db 2644543 CCGCGGCGCGCGGACACGCGCGCTGTGGGCACTTCGATCGCGTCAAGGACGAC 2644602
Qy 322 ATCGCATCGCGGTATGCCACCAAGAACGGGTCCCGATGCTCCGACTGTGATCGCC 381
Db 2644603 GTGACGTTGTGGAGTGCACACCGCTTCGGCACCCAGG---CTATGTCCGCGCTGCT 2644659
Qy 382 ACCGAGGATGCGTGGTGGTGGAGCGGCTGTCTCGCGGAGGCGGCACCATGTCGGCAAG 441
Db 2644660 ACCGACGACTGTAGTGTCTCGCGCGCTCAAGGCGCGGAGCGGTGATCTCGGCAAG 2644719
Qy 442 AGAACCTCGAGGACATGG---CGATGGGTATCGGTGAAGGAGCGGTCTACGGTCTTGG 498
Db 2644720 ACGAATCTTGTGAATTGGGCGAGTGGCGCTTCAACGAGGAGCCCGGTTTCGACACACC 2644779
Qy 499 CTGAACCCGCAACACCCGCCACGCGGTGATCTTCAGCGGCTCCGCGCTGCCGCTGCC 558
Db 2644780 CGCAACCCCTGTCTGGCGCGGACACGCGGGTGGATCTCTGGGCGGTAGCGCGGGCG 2644839
Qy 559 GTGCTGCGCGGCAATGGTTCGACTTTCGCGCTTGGGCGTTCGATGAGCGAGGACGATCCGATC 618
Db 2644840 GTGCGCGCGGCTGTGTACCGCGCTATCGGCTCCGACGCGCGGCGGAGCATCCGATC 2644899
Qy 619 CCGGCGCATGTTGCGGCTGCTGCGGCTGCTGCGGCTGCTGCGGCTGCTTAC 678

Db 2644900 CCGCAGCATGGACACACACCTAGTGGGCATCAAGCCACACACGCGGTGGATCTCCACCTGG 2644959
Qy 679 GGCCTGACATACATGACACACACCTTGGACACATCGGCGCCCATCACACAGGGGGGTGAG 738
Db 2644960 CCGCTCGCGAGCGGTTCACGGGTACCGGTCAACGGGTACTGGCCGCACTGTGGAG 2645019
Qy 739 CTCACGCCCGGGTCTCGAGGTGTGGCCGGGGCGGACGTGGCGGACCTCAGTGGGTG 798
Db 2645020 GATCGCGCGCTGGTGTGACCGCGCTCGGCAACGTCGAGGGCGCGCCACAGCCA 2645079
Qy 799 CGTAACCTTCGGAGCGCGGAGAACTACGGCTCGCGGCTCGG 839
Db 2645080 CCCCCGTGACGGTTCGGATTTCGTGGCAATCGCCCTGG 2645120

RESULT 13

US-08-233-146-8
; Sequence 8, Application US/08233146
; Patent No. 5648256
; GENERAL INFORMATION:
; APPLICANT: BEPPU, TERUHIKO
; APPLICANT: YAMADA, HIDEAKI
; APPLICANT: NAGASAWA, TORU
; APPLICANT: HORINOUCHI, SUEHARU
; APPLICANT: NISHIYAMA, MAKOTO
; TITLE OF INVENTION: A GENE ENCODING A POLYPEPTIDE HAVING
; TITLE OF INVENTION: NITRILE HYDRAZASE ACTIVITY, A TRANSFORMANT CONTAINING THE
; TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
; TITLE OF INVENTION: TRANSFORMANT
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,146
; FILING DATE: 22-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/694,746
; FILING DATE: 02-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: MISROCK, S. LESLIE
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7005-024-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Pseudomonas chlororaphis
; STRAIN: B23 (FERM BP-187)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1021..1620
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1666..2325

RESULT 14

US-08-463-470-8

US-08-233-146-8
Query Match 7.6%; Score 120; DB 1; Length 2456;
Best Local Similarity 47.9%; Pred. No. 1e-14;
Matches 444; Conservative 0; Mismatches 470; Indels 12; Gaps 3;
Qy 614 GGATCCCGCGCATGGTGGGACTGGTGGGATGAAGCGACCCACGCGCTGGTGGCGGT 673
Db 1 GGATCCCGCTCGGCTTCTCGGGTACCTACGGCATGAAGCCACCCACGGCTGGTGGCCT 60
Qy 674 CTTACGGCCTTCACATACATGGACCAACACCTTGGACCACTCGGGCCCATCACAGGGGG 733
Db 61 ACACCGCGCTCATGGGATTGAAGCCACGATCATGATGTCGCCCCCATCACCGGTAACG 120
Qy 734 TCGAGCTCAAGGCCCGGCTCTCGAGGTGTTCGCGGGCCGAGCTGGCGGACCTCAGT 793
Db 121 TGGCGGACNAACGGCGCTGATCTCGAGCAATGGCCGGTGCAGACGACTCGACCCCGGCC 180
Qy 794 GGGTGCCTAACTTCGGAGCCCGAGAACTACCGGCTCCGGCTCGCGAGGGAGTATCCG 853
Db 181 AGCG-----GGCCTCAGGTGATGACTATTTCAGTTACTTGGAAAGGCGTGAGCG 234
Qy 854 GTCTGAGATTGGCGCTCGTCGAGGAGTCACTGGAGCGGAACGGTGCAGCCCGGACGTGA 913
Db 235 GACTCAGATCGGGGTGTTCGAAGAGGGATTTCGGCTTGCTAACCAAGGACCTCGCGTGG 294
Qy 914 TCGCGCGGTTCAACACGAGGACTGGCGCGCTCGAGAGCGCGGTGCGACCATCGAGCGGG 973
Db 295 CGGCAAGTGGCGGAGCGCATCGCCGACTCGAGGGGTGGGCGCTCATGTGAGCGCG 354
Qy 974 TCTCGGTGCCGTGTGAGACGGCGGCTTGCGCTATTCAGAGCGCGGTGATGGCTTTCAACG 1033
Db 355 TCTCAATTCGCGAGCACAACTGGCAGGGTGTGTGGCACCCCATCGGTTGGAGAGGCT 414
Qy 1034 CGCGCGCTATGGCGGACTCTCGCGGTGTGGGCTACTTCCACAAGGGGCGGTGACGCTCA 1093
Db 415 TGACCATGCAAGATGATGATGCAATGGCAACGGCGCAGGCTTTAATCGAAGAGACTTTACGATG 474
Qy 1094 GCACCGCGCTCAGCAGCGCGGCCAGAGTGCACACCCACCAAGGATCTGGCGATCTGT 1153
Db 475 TCGGCTGTGGACAAACAAAGCAGCTGGCGCAGCAGCAGACCAATATTCGCGGTGCG 534
Qy 1154 CCGGCTGTGCTGATCGCGGAGCACCTCGCGCAGCAATACCTCGGCTACCTACCTACG 1213
Db 535 TCAAGCTCTGCTGCTGCTGGCCCATACGCGCTGTGCGCTACACGAGCGCTACTACG 594
Qy 1214 CGAAGCGCGAGAACCTCGCGCTGGAGCTCGGCAAGCAGATCGACGCGCTCTCCAGGACC 1273
Db 595 CCAAGGCCAGAACCTTGCACGCTTTCGCCGCGAGGATACGACAAAGCGCTGCMAACCT 654
Qy 1274 GGGCTGCACTGCTGACCCCGACACCGCTACCGTTGCCAAGAGCTGTGGAGGGTGGC 1333
Db 655 ATGACCTGCTGCTGATGCCACCGCCCATCGCGGCCCAACCCCGCCAG---CGA 711
Qy 1334 AAGACACCATGTCCATGATCCACGATGACGGGCAATGCGATCCTCAACACGCTGCCCGC 1393
Db 712 ACTGCTGATCAGGAGTACGTGGCTCGCGGTGGAAATGATCGGCAATACCGCGCCAC 771
Qy 1394 TGGACCTCAGCGTCAACCGCGCTGACGGTGCACCGGTGCCCGGCGGAGAGGGCTGCG 1453
Db 772 AGGACATCACCGGGCATCCGGCCCATGTGCAATTC---CGTGTGGCTGCTGAGCGGCTGC 828
Qy 1454 CCGTTGGCTCCAAAGTATAGGCGCCACTTCGAGGAGTGCAGCGCTCTACCGACCGCGCG 1513
Db 829 CCGTCGGCTGATGCTGTCGCAAAACACTACGCCGAGGCGACGATTACCAAGCGCGG 888
Qy 1514 CCGTATCGAGGCGCGCGGCTATGG 1539
Db 889 CGGCGTTGAAGCTCGGTGAGCTGG 914

Sequence 8, Application US/08463470
Patent No. 5789211
GENERAL INFORMATION:
APPLICANT: BEPPU, TERUHIKO
APPLICANT: YAMADA, HIDEAKI
APPLICANT: NAGASAWA, TORU
APPLICANT: HORINOBUCHI, SUHARU
APPLICANT: NISHIYAMA, MAKOTO
TITLE OF INVENTION: A GENE ENCODING A POLYPEPTIDE HAVING
TITLE OF INVENTION: NITRILE HYDRATASE ACTIVITY, A TRANSFORMANT CONTAINING THE
TITLE OF INVENTION: GENE AND A PROCESS FOR THE PRODUCTION OF AMIDES USING THE
TITLE OF INVENTION: TRANSFORMANT
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: N.Y.
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/463,470
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/233,146
FILING DATE: 22-APR-1994
APPLICATION NUMBER: US 07/694,746
FILING DATE: 02-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: MISROCK, S. LESLIE
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7005-024-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 750-9090
TELEFAX: (212) 869-9741
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Pseudomonas chlororaphis
STRAIN: B23 (FERM BP-187)
FEATURE:
NAME/KEY: CDS
LOCATION: 1021..1620
FEATURE:
NAME/KEY: CDS
LOCATION: 1666..2325
US-08-463-470-8

Query Match
Best Local Similarity 7.6%; Score 120; DB 1; Length 2456;
Matches 444; Conservative 0; Mismatches 470; Indels 12; Gaps 3;

QY 614 GGATCCCGCGCGATGGTGGGACTGGTGGGATGAAGCGACCCACGCGCTGGTGGCGT 673
Db 1 GGATCCCGTGGCGCTTCTGGGTACCTAGGATGAAGCGACCCACGCGCTGGTGGCGT 60
QY 674 CTTACGGCTGACATACATGGACCAACCTTGGACCAATCGGGGCCATCACAGGGGG 733
Db 61 ACACCGCGGTGATGGCGATTGAAGCCACGATCATGTGCGGCCCATCACCGGTAA 120
QY 734 TCGAGTCAACGCCCGGCTCTGAGGTGTGGCGGGCCGACTGGCGCGACCTCAGT 793

Db 121 TCGCGACAAACGCGCTGATGTGACGCAATGGCCGCTGCAGACGAGCTCGACCCGCGC 180
QY 794 GGGTGGTAACCTTCGGAGCCGGAGAACTACGGCTCCGCGCTCGCGAGGAGTATCG 853
Db 181 AGGCG-----GCGCTCAGGTGCGATGACTATTCAGTTTACCTGGAAGGCGTGAGCG 234
QY 854 GTCTGAGATTCGCGGTGCTCGAGGAGTCACTCGAGCCGGAACCGTGCAGCCGGAAGTGA 913
Db 235 GACTCAGATCGGGGTGTTGCAAGAGGATTCGCGCTTGTCTAACGAGGACCTCGCGTGG 294
QY 914 TCGCCGCTTAACAGGAGTCTGGCGGCTCGAGAGCCGCTGGAGCCGCTGAGCGGG 973
Db 295 CGACAAAGTGGCGGACGCGCATCGCCGCTCGAGCGGTGGGCGCTCATGTGAGCGCG 354
QY 974 TCTCGGTGCGGTGTGCGAGCGGCGCTGCGCTTATCCAGAGCGCGCTGATGGCTTTCAACG 1033
Db 355 TCTCCATTCCGAGACCAACTGGCAGGGTGTGTGGCACCCCATCGGTTCGGAAGGCT 414
QY 1034 CGCGCTATGGCGGACTCGCGGCTGTGGGCTACTTTCACAGGGGCGCTGGAGTCA 1093
Db 415 TGACCATGCGAGATGATGCAATGGCAACGCGCGCAGGCTTTAACTGGAAGACTTTACGATG 474
QY 1094 GCACCGCGCTCACGAGCGGCGCCAGAGTCGCACACCCACAGGATCTGGCGATCCTGT 1153
Db 475 TCGGCTGTGTCGACAAACAGCCAGCTGGCGGACGACGACACCAATTCGCGCTCGC 534
QY 1154 CCGGCTGATGTGTGTGATCGGAGACCTGCGGACGAACTCTCGGATTCACACTG 1213
Db 535 TCAAGCTCTCATGTTCTGTCGCGCAATACGCGCTGTGCGGCTACAAACGAGCGTACTAG 594
QY 1214 CGAAGGCGCAGACCTCGCGCTGGAGCTCGGCAAGCAGATCGAGCGCGCTCTCCAGGAC 1273
Db 595 CCAAGGCCAGAACCTTGCAGCTTTCCGCGCAGGAGTACGACAAAGCGCTGCAACCT 654
QY 1274 GGGTGCATGTGACCCCGACACCCCTACCGTTCCCAACGAGCTGTTGAGCGGTGGC 1333
Db 655 ATGACCTGTGTGTGTCGCGACCCACCGCCCATCACGCGCCCAACCCCGCCAG---CGA 711
QY 1334 AAGACACCATGTCCATGATCCACGGATGACGGCAATCGATCTCTCAACAGCTGCGCG 1393
Db 712 ACTGCTGATCAGGAGTACGTGGCTGCGCGCTTGGAAATGATCGGCAATACCGCGCAC 771
QY 1394 TGGACCTCACCGGTCAACCGCGCTGACGCTGCCACCGGTGCGCGGCGGAGAGGCTGC 1453
Db 772 AGGACATCACCGGCGATCCGCGCATGTCGATTC---CGTGTGCGCTGTGGAGCGGCTGC 828
QY 1454 CGTTGGCTCCAGTGTAGGCGCGCATTCGAGGAGTTCGAGGAGTTCGAGGAGTTCGAGGAG 1513
Db 829 CCGTCGGGCTGATGCTGTGTCGCAAAACATACGCGGAGGCGACGATTTACCAAGCGCG 888
QY 1514 CCGTGTGATCGAGCGCGCGCGCTATGG 1539
Db 889 CGCGTTTGAAGCTCGTGGACTGG 914

RESULT 15
US-09-252-991A-15844
Sequence 15844, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 15844
LENGTH: 1335

TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15844

Query Match 7.3%; Score 114.6; DB 4; Length 1335;
Best Local Similarity 52.7%; Pred. No. 1.1e-13;
Matches 280; Conservative 0; Mismatches 239; Indels 12; Gaps 1;

QY	258	CTGCCGGGTGAGGGGCCACAGAGGGGCCACTGAGCGACCTGACCGCGGATCAAGGA	317
Db	154	CGCCCGGGCGGTGGCGAGCCCTCGGTCCGTGAGCGCATTCCTACACGCCAAGGA	213
QY	318	CTGCATCGCCATCGCGGTATGCCACACAGAGCGGTCCCGATGCTCCGACTGTGAT	377
Db	214	CAGTACCTTGTCAAGGGCTTACCGCGCTTCGCGAGCCCGCGTTCAGGACCTGGT	273
QY	378	CGCCACCGAGGATGCGTGTGTGTGAGCGGTGTTCGCGCAGCGCCACCATCGTCGG	437
Db	274	GGCCACAGCGGATGCTTACCGTTCGAGCGCTTCGCGCGCGCGCGGCGATCTGCTGGG	333
QY	438	CAAGCGAACCTCGAGGACATGGCATGGGTATCGGTGAGGAGCGGTCTACGCTCTGC	497
Db	334	CAAGACCAATGCCGCCCATGSCCAACCGCGGCATGACGCGCGGTCTACGCGCGGC	393
QY	498	GCTGAAACCCGAAACAACCCGCCACCGGCACGGG-----TGGATCTTCCAGCGG	545
Db	394	GGAGAGCCCGTACATATGCCCTTACCTCACCGCGCCCTTCGCTCGGGTCTCCCAATGG	453
QY	546	CTCGGGGCTGCGTCTGCTCGCGGATGTCGACTTCGCCCTTGGGGTCTGAGGCGAGG	605
Db	454	CGCCGACCGCCACCGCGGCCAGCTTCGCCCTTCGGCTTGGCGGAGAAACCTGGTC	513
QY	606	CAGCATCCGGATCCCGCGCGCATGTGCGGACTGTGCGCATGAAGCGACCCACGGCT	665
Db	514	GACCGGGCGCGCGCGCGTGAACAACGGCTGTGCGCTACACCCCTTCGCGCGGGT	573
QY	666	GGTGGCTTTAGCGCTGACATACATGACACACCTTGGACACACATCGGGCCATCAC	725
Db	574	GATCTCGGTGCGCGCAACTGGCCGCTGACCGCGACCATGAGCGTGGTGGTCCCTATGC	633
QY	726	CAGGGGGGTGAGCTCAACGCCCGGGTCTCGAGGTGTTGGCGGGCGGA	776
Db	634	GCGGAGCATGGCCGACCTGCTGGAATCTCTGACGTAGTGTGCGCGAGA	684

Search completed: June 27, 2004, 10:03:41
Job time : 168.584 secs

1	1578	100.0	1578	17	US-10-658-691-4	Sequence 4, Appl
2	208	13.2	1503	15	US-10-156-761-2732	Sequence 2732, Ap
3	208	13.2	9035608	15	US-10-156-761-1	Sequence 1, Appl
4	175.2	11.1	1479	13	US-10-282-122A-3553	Sequence 2553, A
5	172	10.9	1482	13	US-10-282-122A-36120	Sequence 26120, A
6	170.4	10.8	1485	13	US-10-282-122A-38642	Sequence 28642, A
7	152.2	9.6	1488	13	US-10-282-122A-14078	Sequence 14078, A
8	151	9.6	1455	9	US-09-815-342-7934	Sequence 7934, Ap
9	151	9.6	1455	13	US-10-282-122A-30488	Sequence 30488, A
10	142.6	9.0	1362	15	US-10-156-761-7248	Sequence 7248, Ap
11	141.4	9.0	1488	13	US-10-282-122A-11356	Sequence 11356, A
12	141.4	9.0	1536	13	US-10-282-122A-15096	Sequence 15096, A
13	140.6	8.9	1449	13	US-10-282-122A-31392	Sequence 31392, A
14	138.8	8.8	1512	15	US-10-156-761-7324	Sequence 7324, Ap

181 CGACCGCGTACACGACCGCGACATCGCGCGCGAACCGACCGCGGACGACCGCGGTC 240
181 CGACCGCGGTACACGACCGCGACATCGCGCGCGAACCGACCGCGGACGACCGCGGTC 240
241 AAGCGATTATCATCGGTTCTCGCGGTTGAGAGGCGCCACCGAGAGGCGCACTGACGACCTG 300
241 AAGCGATTATCATCGGTTCTCGCGGTTGAGAGGCGCCACCGAGAGGCGCACTGACGACCTG 300
301 ACCGCGCGCATCAAGGACTGCAATCGCATCGCGGTTATGCCACACGAGAAAGGCTCCGG 360
301 ACCGCGCGCATCAAGGACTGCAATCGCATCGCGGTTATGCCACACGAGAAAGGCTCCGG 360
361 ATGCTCCGACATGTGATGCGCAACCGAGGATGCGGTGCTGAGAGGCTGCTCCGCGCA 420
361 ATGCTCCGACATGTGATGCGCAACCGAGGATGCGGTGCTGAGAGGCTGCTCCGCGCA 420
421 GCGCGCACCATCTCGCGCAAGAGCAACCTCGAGGACATGCGGATGCGGTATCGGTGAGGC 480
421 GCGCGCACCATCTCGCGCAAGAGCAACCTCGAGGACATGCGGATGCGGTATCGGTGAGGC 480
481 AGCGTCTACGGTCTCGCTGAAACCGGAAACACCGCGCCACCGGACGCGGTGATCTTC 540
481 AGCGTCTACGGTCTCGCTGAAACCGGAAACACCGCGCCACCGGACGCGGTGATCTTC 540
541 AGCGGTCGCGGCTCGCGTCTCGCGGATGCTGCGGATGCTGCGGCTGCGGCTCGATGAG 600
541 AGCGGTCGCGGCTCGCGTCTCGCGGATGCTGCGGATGCTGCGGCTGCGGCTCGATGAG 600
601 GCAGGACGATCCGCGATCCGCGCGGATGCTGCGGATGCTGCGGATGCTGCGGATGCTGCGG 660
601 GCAGGACGATCCGCGATCCGCGCGGATGCTGCGGATGCTGCGGATGCTGCGGATGCTGCGG 660
661 GCGCTGCTGCGGCTCGCGGATGCTGCGGATGCTGCGGATGCTGCGGATGCTGCGGATGCTGCGG 720
661 GCGCTGCTGCGGCTCGCGGATGCTGCGGATGCTGCGGATGCTGCGGATGCTGCGGATGCTGCGG 720
721 ATCAGGAGGCGGCTCGAGCTCAACCGCGGCTCGCGGATGCTGCGGATGCTGCGGATGCTGCGG 780
721 ATCAGGAGGCGGCTCGAGCTCAACCGCGGCTCGCGGATGCTGCGGATGCTGCGGATGCTGCGG 780
781 CGGACCTCTAGTGGGTGCGTAACCTTCGAGAGCGGAGAACTACGCGCTCGCGCTCGGC 840
781 CGGACCTCTAGTGGGTGCGTAACCTTCGAGAGCGGAGAACTACGCGCTCGCGCTCGGC 840
841 GAGGAGTATCCGCTCTGAGATTCGCGGCTCGCGGATGCTGCGGATGCTGCGGATGCTGCGG 900
841 GAGGAGTATCCGCTCTGAGATTCGCGGCTCGCGGATGCTGCGGATGCTGCGGATGCTGCGG 900
901 ACGCCGAGCTGATCGCGGCTTCAACAGGAGCTCGCGGCTCGAGAGCGCGGCTCGG 960
901 ACGCCGAGCTGATCGCGGCTTCAACAGGAGCTCGCGGCTCGAGAGCGCGGCTCGG 960
961 ACCATCGAGCGGCTCGCGGCTTCAACAGGAGCTCGCGGCTCGAGAGCGCGGCTCGG 1020
961 ACCATCGAGCGGCTCGCGGCTTCAACAGGAGCTCGCGGCTCGAGAGCGCGGCTCGG 1020
1021 ATGCGCTTCAACCGCGGCTATGCGGATTCGCGGCTCGCGGCTTCAACAGGAGG 1080
1021 ATGCGCTTCAACCGCGGCTATGCGGATTCGCGGCTCGCGGCTTCAACAGGAGG 1080
1081 CGCGTGAGCTGAGCAACCGCGGCTACGAGCGGCGGCGGAGTGCACCAACAGGAT 1140
1081 CGCGTGAGCTGAGCAACCGCGGCTACGAGCGGCGGCGGAGTGCACCAACAGGAT 1140
1141 CTGCGGATCTGCTCCGCGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1200
1141 CTGCGGATCTGCTCCGCGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1200
1201 GCGATCCATCTACGAGAGGCGGAGAACTCGCGCTGAGAGTTCGCGAGAGTTCGCGG 1260
1201 GCGATCCATCTACGAGAGGCGGAGAACTCGCGCTGAGAGTTCGCGAGAGTTCGCGG 1260
1261 GTCCTCCAGGACCGCGGCTGCACTGCTGACCCCGGACCGCTACCGCTGCTGCTGCTGCTGCTG 1320

1261 GTCTCTCAGACCGGCTGCACTGCTGACCCCGGACCGCTACCGTTGCAACGAGCTG 1320
1321 TTGAGCGGTGCGGAGACACCATGTCATGATGTCATGATGTCATGATGTCATGATGTCATGTC 1380
1321 TTGAGCGGTGCGGAGACACCATGTCATGATGTCATGATGTCATGATGTCATGATGTCATGTC 1380
1381 AACAGCTGCGCGGTGAGACCTCACCGGTACCCCGCGCTGACGGTGCACCGGTGCGGGC 1440
1381 AACAGCTGCGCGGTGAGACCTCACCGGTACCCCGCGCTGACGGTGCACCGGTGCGGGC 1440
1441 GAGAGGCGCTGCGCGGTGCGCTCCAAAGTATAGGCGCGCACTTCGAGGAGTGCAGCTC 1500
1441 GAGAGGCGCTGCGCGGTGCGCTCCAAAGTATAGGCGCGCACTTCGAGGAGTGCAGCTC 1500
1501 TACCGCACCGCGCGGTGATCGAGGCGCGCGCTATGAGGCTCGCGCGGAGCGGAGC 1560
1501 TACCGCACCGCGCGGTGATCGAGGCGCGCGCTATGAGGCTCGCGCGGAGCGGAGC 1560
1561 GCACCGGTGCTGCGGTAG 1578
1561 GCACCGGTGCTGCGGTAG 1578

RESULT 2

US-10-156-761-2732
; Sequence 2732, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2732
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1503)
US-10-156-761-2732

Query Match 13.2%; Score 208; DB 15; Length 1503;
Best Local Similarity 50.1%; Pred. No. 6.6e-4;
Matches 641; Conservative 0; Mismatches 615; Indels 24; Gaps 4;

QY 262 CGGTGAGGCGCGCACGAGGCGGCACTGAGCGACCTGACCGCGCGGATCAAGGACTGC 321
DB 199 CGGAGCGCGGAGAGCTCGCGCGCTGCGCGCTCGCGCTCGCGCTCAAGACATC 258
QY 322 ATCCGATTCGCGGTATGCCCCACACGAA CGGTTCCCGGATGCTCCGATGTCGCTCGCC 381
DB 259 TTCACACGAGGCGATCCCGACACCGTTCGAGTTCGAGGCTCGGATCGCG 318
QY 382 ACCGAGGATCGGTGAGCGGCTGCTCGCGGCGAGCGGCGCACCATCGTCGGCAAG 441
DB 319 CCTACGAGCGGACCCCTCACCAAGCTGAGCGCGCGGCGGCTGCTCTCGGCAAG 378
QY 442 ACGAACCTCGAGGACATGCGGATGGG---TATCGGTGAGAGCGAGCGTTCGCTCGG 498
DB 379 ACCAACATGAGCGAGTTCGCGCATGGGTCTCTCCACCGAGAACAGCGGTACGCGCC 438


```
Db 3371601 --CGAGCCCGGTCGCGCGGTCGTGAGAGCGCGCCGCAACGCGACGCTCGAGGCGCATG 3371544
Qy 859 AGATTCCGGCTGCTCGAGGAGTCACTTGAGAGCCGAAAGGTCGACGCGGACGCTGATCGCC 918
Db 3371543 CGGTCGGCTGCTCAAGCAATTCGCGGCGAGGGCTACAGGCGGTCGCTCCAGCGC 3371484
Qy 919 GCGTTCAACAGGAGTGGCGGCGCTCGAGAGCGCGGTGCGACCAATCGAGCGGGTCTCG 978
Db 3371483 TTGACAGAGTCCGTCGCTCTCAAGAGAGTGGCGCGGAGATCGTCGAGCTGAGTGC 3371424
Qy 979 GTGCGGTTGAGAGCGGCGCTGCGCTTATCCAGAGCGCGTGTATGCTTTCAACGCGCGC 1038
Db 3371423 CGGCTTCGACCTGCGCTGTGCGGTATCTATGATCGCGCGTCCGAGTCTCTCC 3371364
Qy 1039 GCTATGCGGAGTCCGCGGCTGTGGGTACTTCCACAAGGGGCGGTGAGAGTCAAGC 1098
Db 3371363 AACCTCGCGGCTTCGAGCGGCTGCGTATGCGGTGCGGTGAGTCAAGCGGAGCGCG 3371304
Qy 1099 GCGTCAAGAGCGGCGGCGGAGTGCACCAACCAAGGATCTGCGGATCTCTGCTCCCG 1158
Db 3371303 TCCGCGAGAGGTACCGCGCTCACCGGAGCGGCTTCGCGAGAGGTCAAGCGC 3371244
Qy 1159 CTGATGCTGATGTCGCGAGCACTCGCGAGCAATACCTCGGCATTCCTACTACGCAAG 1218
Db 3371243 CGCATCATGCTCGGACGTACGCGCTCAGCTCGGCTACTACGACGCTACTACGCGAGC 3371184
Qy 1219 GCGCAGAACCTGCGGCTGAGCTCGGCAAGCAGATCGACCGCTCTCCAGGACCGGCT 1278
Db 3371183 GCCAGAGGTCCGCTGCTCATACCGGCACTTCGAGAGTGTTCGAGCAGGTGGAT 3371124
Qy 1279 GCACTGCTGACCCGACGAGCTTACCGTTGCGCAAGAGTGTTCGAGCGTTCGCGAGAC 1338
Db 3371123 GTGATGCTCTCCGACGAGCCGACCCGCTTCCGATTCGCGAGCGCGCGAGC 3371064
Qy 1339 ACCATGCTCATGATCCACGATGACGGGCAATGCGATCTCAACAGTCCGCGCTGGAC 1398
Db 3371063 CCGATGCGAGTATAC-----TCCGCGACCTGTGCACCATCCCGACCAAC 3371019
Qy 1399 CTCACCGGTACCCGCGGCTGACGGTGCACCGGT---CGGCGCGAGAAAGGCGCTGCC 1455
Db 3371018 CTCGCGGCAACGCGCATGTGCTGCTGCGGTCTCGCGCGAGGACACCTCCCG 3370959
Qy 1456 GTTGGCTCAGATGATAGCGCGCACTTCGAGAGTTCGAGCTCTACCGCAGCGGGCC 1515
Db 3370958 GTGCGCTGAGATCATGCGCGCGCGCTCAAGGACGACAGGCTGTACAGGTTCGCGCC 3370899
Qy 1516 GTGATCGAGGCGCGCGCT 1535
Db 3370898 GCGTCGAGGCGCGCTTCGT 3370879
```

RESULT 4

```
US-10-282-122A-25553
; Sequence 25553, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haelebeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 25553
; TYPE: DNA
; LENGTH: 1479
; ORGANISM: Mycobacterium avium
; US-10-282-122A-25553

Query Match 11.1%; Score 175.2; DB 13; Length 1479;
Best Local Similarity 49.6%; Pred. No. 3.9e-32;
Matches 636; Conservative 0; Mismatches 613; Indels 33; Gaps 6;

Qy 285 GGCATGAGCGACCTGACCGCGGATCAAGGACTGCAAGCATCGCCATCGCCGCTATGCCAC 344
Db 213 GCGCTGCGCGGGTGCCTGCGCTGCGCTCAAGAGCGTCTTACACCGTGCATGCCAC 272
Qy 345 CACGAAAGCGTCCGATGCTCCGACTGTGATCGCCACCGAGGATCGCTGGTGGTGA 404
Db 273 CACCTGCGGGTCAAGATTTGCAAGGCTGCGCTGCGCTACGACGCGACCGTCAAC 332
Qy 405 GGGGTGCTCGCGGAGCGCGCCACCATGTCGGCAAGACGAACCTCGAGGACATGGCGAT 464
Db 333 GAAGCTGCGCGCGCGGCGATCCGATCTCGGCAAGACCAACATGAGAGTTCGCGAT 392
Qy 465 GGGTATC--GGTGAAGAGAGCGTCTACGGTCTGCGGTGAAACCCGAAACCCGCCCA 521
Db 393 GGGCTCTCGACCGAGAACTCGGCTACGGCGCGCAACCGCTGGAATGTCGAGCG 452
Qy 522 CGGCAAGGTCGATCTTCCAGCGGCTCGCGGCTGCGCTGCGCTGCGGATGTCGACTT 581
Db 453 GGTGCGCGGTTTCCGCGCGGCGAGCGCGCGCGCTGCGCGGTTCCAGGCGCGCT 512
Qy 582 CGCCTGCGGCTCGATGAGGAGGAGGATCGGATCCCGGCGCGATCGTGGGACTGGT 641
Db 513 GCGGATCGGCTCCGACACCGGCGGCTCGATCCGCGAGCGCGCGGCTGACCGGACCGT 572
Qy 642 CGGATGAAGGAGCGACCGCGGCTGCTGCGCTTTCAGCGCTGACATACATGAGGACAC 701
Db 573 CGGGGTCAAGCCCACTACGGCACCGGTGTCGCGTACGCGCTGCTGGGCTGCGCGCTATC 632
Qy 702 CTGGAACACATCGGCGGCGCATCACGAGGGGGTTCGAGTCAACGCGCGGCTCTCGAGT 761
Db 633 GCTGGATCAGGGGCGGCGGCTGCGCGGCGCACCGTTCGACACCGCGCTGTCACAGGT 692
Qy 762 GTTGGCGGCGCGGAGCTGCGGCGGACCTCAGTGGTGGTAACTTCGCGGCGCGGAGNA 821
Db 693 GATCGCGGCGGACGACATCCGCGACTCCACCTCCGT---CGACGCGCGGTCGCGCGT 749
Qy 822 CTACGGCTCGCGCTCGGCGAGGAGTATCCGCTCTGAGATTCGCGGCTCGTGGAGGTC 881
Db 750 CGTGGCGCGCGCGCGCGCGCGCGCGCGGTCGAGTCAAGGGTGTTCGCGTGGGGTGGT 809
```

Qy 882 ACTGAGCCGACCGGTGCG-----ACGCCGAGCGTATCGCCCGTTCACACGAGGACT 935
Db |||||
Qy 810 CAAGCAGCTGCGCGCAGGAGATATCAGCCCGCGGTGCTGCGCTCTCTTCGAGCGCGCGT 869
Db |||||
Qy 936 GCGCGCGCT---CGAGAGCGCGGTGCGACATCGAGCGGTCTCGGTGCGGTGCGAC 992
Db |||||
Qy 870 CGAGAGTTGACCGCGCGCGGCGCGAGGTGAGCGGTGAGGTGAGTCCCGGACATTCGAGTA 929
Db |||||
Qy 993 GCGCGCGCTATCCAGAGCGCGGTGAGGCTTTCAAAGCGCGCGCTATGCGGACTC 1052
Db |||||
Qy 930 GCGCGCGCGCTACTACCTGATTTTGCCTCGGAGGTGCCAGCAACCTGGCGCGCTT 989
Db |||||
Qy 1053 GCGCGGTGCGGTACTTCCACAGGCGCGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT 1112
Db |||||
Qy 990 GAGCGGATGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 1049
Db |||||
Qy 1113 GCGCGAGTGCACACCAACAGGATCTGCGGATCTGTCGCGGTGATGCTGTGAT 1172
Db |||||
Qy 1050 GATGCGGTGCGCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 1109
Db |||||
Qy 1173 GCGGAGACCTGCGCGGAGATACCTCGGATCCACTACGAGGCGGAGGAGGAGGAGGAGG 1232
Db |||||
Qy 1110 CACCTACGCTGTCGCGGCTACTACGAGCGCTACTACACGAGGCGGAGGAGGAGGAGG 1169
Db |||||
Qy 1233 GCTGAGTGCAGAGATGCGAGCGGTCTCCAGGAGCGGTGCGGTGCGGTGCGGTGCGGT 1292
Db |||||
Qy 1170 CACCTGATGCGCGGATCTCGAGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGT 1229
Db |||||
Qy 1293 GACGAGCTGCGGTGCGGAGGAGTCTGAGCGGTGCGGAGGAGGAGGAGGAGGAGGAGG 1352
Db |||||
Qy 1230 GCGCACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1289
Db |||||
Qy 1353 CCCAGGATGAGGCGCAATGCGATCTCAACAGGTCGCGGTGCGGTGCGGTGCGGTGCGGT 1412
Db |||||
Qy 1290 TCTGTGCA-----CCTGTGACGCTGCGGTGCGGTGCGGTGCGGTGCGGT 1334
Db |||||
Qy 1413 GCGGTGAGGTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1469
Db |||||
Qy 1335 GCGCATGTCGCGGATCGGATGTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1394
Db |||||
Qy 1470 GATGAGCGGCGGATGCGAGGAGTGCAGCTCTACCGAGGAGGAGGAGGAGGAGGAGGAG 1529
Db |||||
Qy 1395 CATGCGCGCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1454
Db |||||
Qy 1530 GCGGCTATGAGGAGTCTGCGCGG 1551
Db |||||
Qy 1455 GCGGCGCGGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1476
Db |||||

RESULT 5

US-10-282-122A-26120
; Sequence 26120, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Cart, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Foreyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 26120
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
US-10-282-122A-26120

Query Match 10.9%; Score 172; DB 13; Length 1482;
Best Local Similarity 49.0%; Pred. No. 2,2e-31;
Matches 624; Conservative 0; Mismatches 620; Indels 30; Gaps 5;

Qy 275 CCACGAGGCGGACCTAGCGACCTGACCGCGCGGATCAAGGAGTGCATCGCATCGCG 334
Db CCTTCGCTCGCGCTGCGCGGCTGCGCGGCTCAAGGAGCTTCAACACGAGCG 282

Qy 335 GTATGCCACCAAGACGCGTCCGATGCTCCGACTGATCGCATCGCACCGAGGATCGCG 394
Db ACATGCCACCACTCGCGGTCAAAATCTTGAGGAGTGGCATCTCTACGAGCGCA 322

Qy 395 TGGTGTGAGGCGGTCTCGCGGAGGCGCCACCACTCTCGGCAAGAGCAACCTCGAGG 454
Db CGCTGACCGCGGTTGCGCGCGGCGGATCCGATCTCGGCAAGAGCAACATGAGACG 382

Qy 455 ACATGCGGATGG---TATCGGTGAGGAGGAGTCTACGCTCTCGGTGAGACCGGAC 511
Db AGTTCGAGTGGGCTCGTCAAGGAGTCTCGGCTTACGCTTCCACCGGCAACCGGTGA 442

Qy 512 ACCCGCGCCACGACGCGGTGGATCTTCCAGCGGCTCCGCGCTCGCTCGCTCGCGGCA 571
Db ATCTGACCGGTACCGCGGCTTCCGCTGGGCGGAGCGCGGCGGCTGCGCGCTTC 502

Qy 572 TGGTGCATCTCGCGCTGGGCTCGATGAGGAGGAGGATCCGAGTCCCGGCGCATGGT 631
Db AGGCGCGCTGGGCTCGGATCCGATCCGCGGCGGTGATCCGCGGAGCGCGCGCTGA 562

Qy 632 GCGACTGTGTCGATGAAGGCGACCAACCGGCTGCTGCGCTCTTACGCGCTGACATACA 691
Db CCGGACCGCTCGGCTGCAACACCACTACGCGGAGTGTGCGCTATGGCTGTGGCT 622

Qy 692 TGGACCAACCTTGGACCAATCGGCGCCATCAACGAGGAGGAGTCAAGCTCAAGCGCGG 751
Db GCGGCTCTCGCTGATGAGGCGGCGGCTGCGGCGGAGCGGCTTCTTGAACCGCGCTGT 682

Qy 752 TCCTCGAGGTGTTGGCGGCGGCGGAGTGGCGGCGGAGCTCACTGAGTGGTGGTAACTTC 811
Db TGCATGAGGTGATCGCGGCGGAGTGGCGGCGGAGTGGCGGAGTCTGCGGTGCGGCTCG 742

Qy 812 AGCGGAGAACTACGGCTCCGCTCGGCGGAGGAGTATCCGCTCTGAGATTCGCGGTGCG 871
Db CCGAGGTGGTGGCGCGCTAGGCGGCGGCTGCGGAGTCTGCGGTGCGGCTGCGGCTCG 802

Qy 872 TCGAGGAGTCACTGAGGCGGAGCGGTCG-----ACGCGGAGGAGTCACTCGCGGTCA 925
Db CGGTGTTGACAGCTGCAAGGCGGCGGAGGAGTCAACGCGGCGGCTGCTGCGCTCTCTTCG 862

QY 926 ACCAGGAGTGGCGGCTCGAGAGGCGGCTCGACCATCGAGCGGCTTCGGTGCCT 985
 Db 863 AGGCTGCTGGAGGAGCTAACCGGCTGGGCGCTGAGGTACGAGGTGCGACTGCCCGC 922
 QY 986 TGTGACGCGGCTGGGCTATCCAGAGCGGCTGATGGCTTTCACG---CGCGCGTA 1042
 Db 923 ACTTCGACCATGCCCTGGCGGCTATTACTGATCTGCCCTCGAGGTGTCGACATC 982
 QY 1043 TGGCGAGCTCGCGGCTGGGCTATTCACAAAGGGGCGGTGAGCTGACGACCGCG 1102
 Db 983 TGGCGGCTTCGACGCGATGCGCTACGGGCTCGCGGCTCGCGACGACGCGACCGCG 1042
 QY 1103 TCACGACGCGGCGGCGGCTCGACACCCACACAGATCTGCGGCTCTGTCGCGGCTGA 1162
 Db 1043 CCGAGAGGTGATGGCGATGACCGGCGCGCGGTTTCGGGCTCGAGGTCAAGCGGCG 1102
 QY 1163 TGTGCTGATTCGCGGAGCACCTGCGGAGCAATACCTGCGCATCCACTACGCGAAGCGC 1222
 Db 1103 TCATGATCGGCACTACGGGCTTTCGGCGGCTACTACGACGCTATTACACACGCGC 1162
 QY 1223 AGAAGCTCGGCTGGAGCTCGGACAGCATCGAGCGCTCTCCAGACCGGCTGCAC 1282
 Db 1163 AGAGGTGCGACGCTGATCGCGGACCTCGACCGGCTATCGTCCGTGACGTGC 1222
 QY 1283 TGTGACCGGCGGCTGACGCTACGCTTGCACAGAGCTGTGAGCGGTGCGGACGACCA 1342
 Db 1223 TGTGTCGCGGCACTACGCGGCTTCCGGCTGGGTGAGAGGTGACGATCGC 1282
 QY 1343 TGTCCATGATCCACCGATGACGCGGCAATCGCATCTCAACGCTGCGGCTGACCTCA 1402
 Db 1283 TGGCGATGATCTGTTGCA-----CCTGTGACGCTGCGGCTGAACTGG 1327
 QY 1403 CCGGTCAACCGGCGCTGACGCTGCCACGCGTG---CGGCGAGAGGCGCTTGCCTGTG 1459
 Db 1328 CCGGCACTGCGGCTATGTCGCTGCGGCTTCCCGGACGACGCGTGTGCGGTTG 1387
 QY 1460 GCTTCAAGTATAGCGCGGCTTTCGAGAGTTCGACGCTTACGCGACCGCGCGGTGA 1519
 Db 1388 GCTTACAGATCATGCGCGGCAATGGCGGCAATGGCGGCTTACCGGCTTACCGGCTT 1447
 QY 1520 TCAGGCGCGCGC 1533
 Db 1448 ATGAGGCGCGCGC 1461

RESULT 6

US-10-282-122A-28642
 ; Sequence 28642, Application US/10282122A
 ; Publication No. US20040029129A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26

QY 275 CCACGAGGCGGCTGAGCGACCTGACCGCGGCTCAAGGACTGATCGCCATCGCGC 334
 Db 203 CTTGCGGCTGGCGGCTGGCGGCTGGCGGCTCAAGGACGCTTCAACCAAGCG 262
 QY 335 GTATGCCACACGAGCGGCTCGGATGCTCCGACTGATCGCCACGAGGATGCCG 394
 Db 263 ACATGCCACACCTGCGGCTCAAAATCTTGAGGAGTGGCGATCTCCCTAGACGCA 322
 QY 395 TGGTGTGAGCGGCTCTCGCGGCGGCGGCGGCTCCGATCTTGAGGAGTGGCG 454
 Db 323 CGCTGACCGGCTGGCGGCGGCGGCGGCTCCGATCTTGAGGAGTGGCG 382
 QY 455 ACATGGCGATGG---TATCGGTGAGGAGCGCTTACGCTCTGCGCTGAACCGGACA 511
 Db 383 AGTTGCGGATGGGCTCGTTCGACGAGAACTCCGCTTACGCTTCCACCGAACCGGTGA 442
 QY 512 ACCCGCGCGGCGGCTGAGTCTTTCAGCGGCTCGCGGCTCGCTCGCTCGCGGCA 571
 Db 443 ATCTGACCGGCTACCGCGGCTTCGCTGCGGCGGCGGCGGCTGGCGGCTTCC 502
 QY 572 TGGTCACTTTCGCTCGGCTCGATGAGGACGAGCATCCGATCCCGGCGGCTCGAT 631
 Db 503 AGGCGCGCTGGCATCGGATCCGACACCGGCGGCTCGATCCGCGGCGGCGGCTGA 562
 QY 632 GCGGACTGCTCGGCTGAGGCGGCGGCGGCGGCTGGTGGCTCTTACGGCTTGACATCA 691
 Db 563 CCGGACCGTGGGCTCAACCCACTTACGCGGAGGTGCTGCGCTATGGGCTTGGTGGCT 622
 QY 692 TGGACCAACACTTGGACCATCGGCGGCTTCAACGAGGCGGCTGAGCTCAACCGCGG 751
 Db 623 GCGGCTCTGCTGATCAGGCGGCGGCTGCGGCGGCGGCTTCTTGACACCGGCGCTGT 682
 QY 752 TCCTCGAGGTGTTGGCGGCGGCGGCTGCGGCGGCGGCTCAGTGGTGGCTGAACCTTCGG 811
 Db 683 TGCATCAGGTGATCGCGGCGGCGGCTGCGGCGGCGGCTCAGTGGTGGCTGAGGAGTGC 742
 QY 812 AGCGGAGAACTACGGCTTCGCGGCTCGGCGGAGGAGTATCCGCTCTGAGATTCGCGGTCG 871
 Db 743 CCGAGCTGGTGGGCGCGCTAGGCGGCGGCTGGGCGGCTGCGGCTGCGGCTGCG 802
 QY 872 TCGAGGAGTCACTGAGCGGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTTCA 925
 Db 803 GCGTGGTTCGACAGCTGACGCGGCGGCGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCT 862
 QY 926 ACCAGGAGTGGCGGCTCGAGAGCGCGGCTGAGGAGCGGCTTTCGAGGAGGCTTTCGAGG 985

QY 926 ACCAGGAGTGGCGGCTCGAGAGGCGGCTCGACCATCGAGCGGCTTCGGTGCCT 985
 Db 863 AGGCTGCTGGAGGAGCTAACCGGCTGGGCGCTGAGGTACGAGGTGCGACTGCCCGC 922
 QY 986 TGTGACGCGGCTGGGCTATCCAGAGCGGCTGATGGCTTTCACG---CGCGCGTA 1042
 Db 923 ACTTCGACCATGCCCTGGCGGCTATTACTGATCTGCCCTCGAGGTGTCGACATC 982
 QY 1043 TGGCGAGCTCGCGGCTGGGCTATTCACAAAGGGGCGGTGAGCTGACGACCGCG 1102
 Db 983 TGGCGGCTTCGACGCGATGCGCTACGGGCTCGCGGCTCGCGACGACGCGACCGCG 1042
 QY 1103 TCACGACGCGGCGGCGGCTCGACACCCACACAGATCTGCGGCTCTGTCGCGGCTGA 1162
 Db 1043 CCGAGAGGTGATGGCGATGACCGGCGCGCGGTTTCGGGCTCGAGGTCAAGCGGCG 1102
 QY 1163 TGTGCTGATTCGCGGAGCACCTGCGGAGCAATACCTGCGCATCCACTACGCGAAGCGC 1222
 Db 1103 TCATGATCGGCACTACGGGCTTTCGGCGGCTACTACGACGCTATTACACACGCGC 1162
 QY 1223 AGAAGCTCGGCTGGAGCTCGGACAGCATCGAGCGCTCTCCAGACCGGCTGCAC 1282
 Db 1163 AGAGGTGCGACGCTGATCGCGGACCTCGACCGGCTATCGTCCGTGACGTGC 1222
 QY 1283 TGTGACCGGCGGCTGACGCTACGCTTGCACAGAGCTGTGAGCGGTGCGGACGACCA 1342
 Db 1223 TGTGTCGCGGCACTACGCGGCTTCCGGCTGGGTGAGAGGTGACGATCGC 1282
 QY 1343 TGTCCATGATCCACCGATGACGCGGCAATCGCATCTCAACGCTGCGGCTGACCTCA 1402
 Db 1283 TGGCGATGATCTGTTGCA-----CCTGTGACGCTGCGGCTGAACTGG 1327
 QY 1403 CCGGTCAACCGGCGCTGACGCTGCCACGCGTG---CGGCGAGAGGCGCTTGCCTGTG 1459
 Db 1328 CCGGCACTGCGGCTATGTCGCTGCGGCTTCCCGGACGACGCGTGTGCGGTTG 1387
 QY 1460 GCTTCAAGTATAGCGCGGCTTTCGAGAGTTCGACGCTTACGCGACCGCGCGGTGA 1519
 Db 1388 GCTTACAGATCATGCGCGGCAATGGCGGCAATGGCGGCTTACCGGCTTACCGGCTT 1447
 QY 1520 TCAGGCGCGCGC 1533
 Db 1448 ATGAGGCGCGCGC 1461

Query Match 10.8%; Score 170.4; DB 13; Length 1485;

Best Local Similarity 48.9%; Pred. No. 5.3e-31;

Matches 623; Conservative 0; Mismatches 621; Indels 30; Gaps 5;

QY 275 CCACGAGGCGGCTGAGCGACCTGACCGCGGCTCAAGGACTGATCGCCATCGCGC 334

Db 203 CTTGCGGCTGGCGGCTGGCGGCTGGCGGCTCAAGGACGCTTCAACCAAGCG 262

QY 335 GTATGCCACACGAGCGGCTCGGATGCTCCGACTGATCGCCACGAGGATGCCG 394

Db 263 ACATGCCACACCTGCGGCTCAAAATCTTGAGGAGTGGCGATCTCCCTAGACGCA 322

QY 395 TGGTGTGAGCGGCTCTCGCGGCGGCGGCGGCTCCGATCTTGAGGAGTGGCG 454

Db 323 CGCTGACCGGCTGGCGGCGGCGGCGGCTCCGATCTTGAGGAGTGGCG 382

QY 455 ACATGGCGATGG---TATCGGTGAGGAGCGCTTACGCTCTGCGCTGAACCGGACA 511

Db 383 AGTTGCGGATGGGCTCGTTCGACGAGAACTCCGCTTACGCTTCCACCGAACCGGTGA 442

QY 512 ACCCGCGCGGCGGCTGAGTCTTTCAGCGGCTCGCGGCTCGCTCGCTCGCGGCA 571

Db 443 ATCTGACCGGCTACCGCGGCTTCGCTGCGGCGGCGGCGGCTGGCGGCTTCC 502

QY 572 TGGTCACTTTCGCTCGGCTCGATGAGGACGAGCATCCGATCCCGGCGGCTCGAT 631

Db 503 AGGCGCGCTGGCATCGGATCCGACACCGGCGGCTCGATCCGCGGCGGCGGCTGA 562

QY 632 GCGGACTGCTCGGCTGAGGCGGCGGCGGCGGCTGGTGGCTCTTACGGCTTGACATCA 691

Db 563 CCGGACCGTGGGCTCAACCCACTTACGCGGAGGTGCTGCGCTATGGGCTTGGTGGCT 622

QY 692 TGGACCAACACTTGGACCATCGGCGGCTTCAACGAGGCGGCTGAGCTCAACCGCGG 751

Db 623 GCGGCTCTGCTGATCAGGCGGCGGCTGCGGCGGCGGCTTCTTGACACCGGCGCTGT 682

QY 752 TCCTCGAGGTGTTGGCGGCGGCGGCTGCGGCGGCGGCTCAGTGGTGGCTGAACCTTCGG 811

Db 683 TGCATCAGGTGATCGCGGCGGCGGCTGCGGCGGCGGCTCAGTGGTGGCTGAGGAGTGC 742

QY 812 AGCGGAGAACTACGGCTTCGCGGCTCGGCGGAGGAGTATCCGCTCTGAGATTCGCGGTCG 871

Db 743 CCGAGCTGGTGGGCGCGCTAGGCGGCGGCTGGGCGGCTGCGGCTGCGGCTGCG 802

QY 872 TCGAGGAGTCACTGAGCGGCGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTTCA 925

Db 803 GCGTGGTTCGACAGCTGACGCGGCGGCGGCTGAGGAGGCTGAGGAGGCTGAGGAGGCT 862

QY 926 ACCAGGAGTGGCGGCTCGAGAGCGCGGCTGAGGAGCGGCTTTCGAGGAGGCTTTCGAGG 985

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: Patentin version 3.1

SEQ ID NO 28642

LENGTH: 1485

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

US-10-282-122A-28642

863 AGGTCGCGTGGAGCAGCTAACCGCGCTGGCGCGTGGAGTCAAGGAGTCAAGTGGCGCG 922
Qy 986 TGTGAGCGCGCGCTGCGCTATCCAGAGCGCGGTGATGGGTTTCAACG---CGCGCGCTA 1042
Db 923 ACTTCGACCATGCGCTGCGCGCTATTACCTGATTTCTGCGCTCGGAGGTCTCGAGCAATC 982
Qy 1043 TGGCGGACTCGCGCGGTGTGGGCTACTTCCACAGGCGCGGTGAGCTCAGACCGCGCG 1102
Db 983 TGGCGCGCTTCAGACGAGTGCCTACCGGGTTCGCGGTGCGGACGACGACCGGAGCG 1042
Qy 1103 TCACAGCGCGCGCCAGAGTCGACACCAACCAAGATCTGGGATCTTCCCGGTGA 1162
Db 1043 CCGAGGAGGTGATGCGGATGACCGCGCGCGCGGTTTCGGGCGCGAGGTCAAGCGCGCA 1102
Qy 1163 TGCTGGTGTATCGCGGAGCACCTGCGGACGAGTACCTCGCATCTCACTACGAGGCGC 1222
Db 1103 TCATGATCGGCACCTACCGGTGTGCGCGCGCTACTAGACGCTATTACACGAGCGC 1162
Qy 1223 AGAACCTTCGCGCTGGAGCTCGGCAAGCAGATCGACGCGCTCTCCAGGACCGGCTGCAC 1282
Db 1163 AGAAGGTGCGCAGCTGATCGCGCGGACCTCGACGCGGCGTATCGGTTCGTCGACGTGC 1222
Qy 1283 TGCTGACCGCGGACAGCGCTACCGTTGCCAAGAGCTGTTCAGCGGTGCGGACGACCA 1342
Db 1223 TGGTGTCCGCCACGACCGCGCGACCGCGTTCCGCGATGGGTGAGAGGTGAGCATCCG 1282
Qy 1343 TGTCCATGATCCACGAGTACGCGCAATGCGATCTCAACACGTCGCGCGTGCACCTCA 1402
Db 1283 TGGCGATGACTTGTTCGA-----CCTGTGACGCTGCGCGTGACTTGG 1327
Qy 1403 CGGTTCACCGCGCGCTGACGGTGCACCGG---TGCGGGCGAGAAGGCGCTGCCCGTTG 1459
Db 1328 CGGCGCACTGCGGCACTGTCTGTGCGGTGCGGCGCTCTCCCGGACGACGCGTTCGCGGTTG 1387
Qy 1460 GCTCCCAAGTATAGCGCGCACTTCGAGGAGTCAAGCTCTACCGCACGCGCGCGCTGA 1519
Db 1388 GCCTACAGATCATGCGCGGCGCATTTGGCGGACGCGGCTCTACCGGTTGGGCGGCTT 1447
Qy 1520 TCGAGCGCGCGCGC 1533
Db 1448 ATGAGCGCGCGCGC 1461

RESULT 7

US-10-282-122A-14078
; Sequence 14078, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14078
; LENGTH: 1488
; TYPE: DNA
; ORGANISM: Burkholderia mallei
; US-10-282-122A-14078

Query Match 9.6%; Score 152.2; DB 13; Length 1488;
Best Local Similarity 48.6%; Pred. No. 1.1e-26;
Matches 631; Conservative 0; Mismatches 623; Indels 45; Gaps 6;
Qy 273 CGCCACGAGGGGCGCTAGAGCGA CTTGACCGCGCGCATCAAGGACTGATCGCCATCGC 332
Db 183 CGCGGAGGCGGGCGCGCTCGCGCGCTCGCGATCGTGCAAGGACGTTTCTGTACACGG 242
Qy 333 CGGTATGCCACACACGAGCGGTCCCGATGCTCCGACTGTGATCGCCACCGAGATGC 392
Db 243 CGGCTGCGCTCCACGCGCGGTTCGAGATGTCGCGAATCTAGCGAGCGCGTTCGACGC 302
Qy 393 CGTGTGTGTGAGCGGCTCTCGCGCAGCGCGCACCATCTGTCGGCAAGACCACTCGA 452
Db 303 GACCGTCTGTCGCGCGCTGTGCGCGCGCGCATGTGTACGCTCGGCAAGCAACATGGA 362
Qy 453 GGCATGCGGATGGGT---ATCGGTGAAGCGAGCGCTACGCTCTGCGTGAACCCGAA 509
Db 363 CGAGTTCGCGATGGGTTCGTGCAACGAGAACTCCGCGTTCGCGCGGTGAGNACCGGTG 422
Qy 510 CAACCGCGCGCCACGCGCAGCGGTGATCTTTCAGCGCGCTCCGCGCTGCGCTCGCTCGCG 569
Db 423 GGACACGAGCGGTGCGCGCGCGGAGCTCGGCGCGGAGCTCGCGCGCGCTCGCGCGCGC 482
Qy 570 CATGCTGACTTCGCGCTGGCGTTCATGAGCGAGGAGCATCCGATCCCGCGCGCATG 629
Db 483 CTTGCGCGCGCGCGCAGCGGCGCACACACGCGCGCTCGATCCGCGAGCGCGCTGTT 542
Qy 630 GTGCGGACTGCTGCGCATGAAGCGCGACCCACCGCTTGGTTCGCTTTACGCGCTGACATA 689
Db 543 CGCGCGGTGACCGGCTCAAGCGCGCTACGCGCGCTGTGCGCTACGGGATGATGCG 602
Qy 690 CATGACACACCTTGGACACATCGGCGCGCATCACAGGGGGGTGAGTCAACCGCGG 749
Db 603 GTTTCGCGTCTGCTCGATCGACCGCGCGCGCGATGGGCGGAGCGCGCGGATTTGCGCGCT 662
Qy 750 GGTCTCTGAGGTGTGTCGCGCGCGCGCTGCGCGCGACCTTCAGTGGGT-----797
Db 663 CTTGCTGAACGCGATGGCGGCTTCGACGAGCGGATTCGACGAGCTCGAGCGCGCGA 722
Qy 798 -----CGTAACTTTCGAGCGCGGAGAACTACGCTTCGCGCGCGCGCGCGCGCGA 845
Db 723 CGAAGACTACACGCGCGCACCTCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCG 782
Qy 846 AGTATCCGCTCGAGATTCGCGGTTCGTGAGGAGTCACTGAGCGCGCGCGCGCGCGCGCG 905
Db 783 GCTCGCGCGCGCTCGCGCATCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839
Qy 906 GGACGTGATCGCGCGCTTCAACAGGAGTTCGCGCGCGCTCGAGAGCGCGCGCGCGCGCAT 965
Db 840 CGACGTGCGCGCGCGATCGACGCGCGCTGAAGAGCTACGAGCGCGCTCGCGCGCGCGCT 899
Qy 966 CGACCGGCTCTCGGTGCGCGCTTGTGGAGCGCGCGCTGCGCTTATCCAGAGCGCGCGTGTGCG 1025

Db 900 CTGCGCGTCTGCTGCCGAGAGCGAGCTGTCTGATTCGCTTACTATGTATGATCGCC 959
Qy 1026 TTTCAACGCGCGGCTA---TGCGGACTCCGCGGTGGGTACTTCCACAAGGGGG 1082
Db 960 CGCGAGGCTCGTGAACCTGTGCGCTTCGACGGGTGCTACGGCCAC-----CG 1013
Qy 1083 CGTGGAGCTCAGACCGCGCTCAGACGGGGCCGAGAGTCGACACCCACAGGATCT 1142
Db 1014 TGGGCGCAATACCGCATCTGCTCGACATGTAAAGAGTCGCGCGGAGGGCTTCGG 1073
Qy 1143 GGCATCTCTCCCGGTGTGTGTGATCGCGGAGCACCTCGCGGAGCAATACCTCGG 1202
Db 1074 GCCTGAGTGAAGCGCGGATCTCGTCGCGACGTAGTGTCTGCGACGGCTACTACGA 1133
Qy 1203 CATCTACTAGCAAGCGCAGACCTCGCGCTGGAGCTCGGCAAGCATCGACCGCT 1262
Db 1134 CGCTACTCTGAGCGCGAGATCCCGTCCAGATGTATC-----TGGCGGATATCTATACGCTGTC 1193
Qy 1263 CCTCCAGGACGGGCTGCACTGTGACCCCGACACGCTTACCGTTTGCCAAACGAGCTGT 1322
Db 1194 GTTCAATCTCGGACGTGATCATGGGCGCGGTGCGCCACCGTGGGACATCGG 1253
Qy 1323 GAGCGTCCGCAAGACACCATGTCCATGATCCACGGATGACGGGCAATGCGATCTCAA 1382
Db 1254 CGGAAAGGGCGAGATCCCGTCCAGATGTATC-----TGGCGGATATCTATACGCTGTC 1307
Qy 1383 CAGTGCGCGCTGACCTCACCGGTACCGCGGCTGACGTTGCCACGCGGTGCGGGCGA 1442
Db 1308 GGTGAGCTCGCGGCTTCCCGGATGAGCTGCGGTGCGGTGCGGGCGGCGGA 1367
Qy 1443 GAGGGCTCGCGGTGCGCTCCAAAGTATGAGCGCGGCTTCCGAGAGTCGACGCTCTA 1502
Db 1368 CGGAAAGCGCGGTGCGGTGCGATCATCGGCAACTATTTCGAGAAAGCGCGGATGCT 1427
Qy 1503 CCGCACCGCGCGTGTATGAGCGCGCGGCGCTTATGGGA 1541
Db 1428 GCAGGTGCGCGAGCGGTTCAGCGCGGACCGAGTGGCA 1466

RESULT 8

US-09-815-242-7924
; Sequence 7924, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7924
; LENGTH: 1455
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1455)
US-09-815-242-7924

Query Match

Best Local Similarity 9.6%; Score 151; DB 9; Length 1455;
Matches 598; Conservative 0; Mismatches 590; Indels 33; Gaps 6;

Qy 273 CGCACGAGGGGCACTGAGCGACCTGACCGCGCGATCAGAGACTGCAATCGCATCGC 332
Db 189 CGCGAGAACGCGCCCTGTCTCGCGCGCGATCGCCCAAGAGACCTGTTCTGCACCCA 248
Qy 333 CGGTATGCCACCAACGAGGTCCTCGGATCTCCGACTGTGATCGCCACCGAGGATGC 392
Db 249 GGGCTAGCGACAGCTCGGTTGGAAGATGCTGACAACTTCTGCTCGCCCTACGACGC 308
Qy 393 CGTGTGTGAGCGGCTGCTCGCGGACGGCGCACCATGCTCGGCAAGAGAACTCGA 452
Db 309 CACGCTGTGAGAGAGCTCACCGCGCGCGGCTTACCTCGGCAAGCTGAATGGA 368
Qy 453 GGAATGCGGATGG---TATCGGTGAAGGAGAGCTGTACGGTCTCTGCGTGAACCCGAA 509
Db 359 TGAATTGCCATGGCTCGTGAACCACTGACGCCATACCGCGGGTGAAGAACCCCTG 428
Qy 510 CAAACCGCGCCACGCGACGGGTGATTTCAAGGGTTCGCGCTCGCGCTGCGGTGCGGG 569
Db 429 GAGCTCAGCGCGTTCGCGCGGCTCTCTCGCGGCTTCGCGCGGCGAGTCGCGCGGCG 488
Qy 570 CATGTCGACTTCGCGCTGCGGTCGATGAGGAGGAGGAGCATCGGATCCCGCGCGCATG 629
Db 489 CTGCTCGCGCGCGCACCGGACCGATACCGCGGCTCGATTCGCCAACCGCGGCGCT 548
Qy 630 GTGCGAATGTCGCGATGAAGGCGACCCACCGGCTGTGCTGCTTACGCGCTGACATA 689
Db 549 GACCAACTCTACCGGATCAAGCCAACTACCGCGGCTTTCGCGCTGGGCGCATGATCGC 608
Qy 690 CATGCGACACACCTTGGACACATCGCGGCCCATCACAGGGGGTTCGAGCTCAAGCGCG 749
Db 609 CTACGCTTCCAGCTTCGACAGGGCGGCGCTCGCGGCGCACCGCGGAGTTCGCGCT 668
Qy 750 GGTCTCGAGGTGTTGGCGGGCGGAGCTGCGCGGACCTCAGTGGGTGCGTAACCTTCC 809
Db 669 GATGCTGGGGTGTGTCGGGATTCGATCCGAAGGACTCGACACAGCT-----CGAACA 722
Qy 810 GGAGCGGAGAACTACGCTCGCGCTCGCGAGGGAGTATCGGTCTGAGATTTCGCGGT 869
Db 723 GCGGTGACGACTACTTGGCGCGCTGCAAGACCGCTGAGCGGCTTCGCGATCGGCT 782
Qy 870 GGTGAGGAGTCACTGGAGCGGAGCGGTGCGACCGGACGCTGATCGCGCGGTTCACCA 929
Db 783 CGCGCGGAATATCTCGCGCGCGCTCGACAGCGGATCGCGCGGCTGCGCGT 842
Qy 930 GGGACTGCGGCTCGAGAGCGCGGTGCGACCATCGAGCGGCTCTCGGTGCGGTG 989
Db 843 G---TCGAGGAGCTGAAGAGCTCGCGCGCACCGTGAAGGACATTTTCTTCGCCAACA 899
Qy 990 GAGCGCGGCTGGGCTATTCAGAGCGGCGTGTATGGCTTTTCAACGCGCGCTATGGCGGA 1049
Db 900 GCAGCACCCATCCCGGCTACTAGTAATCGCGCGCGGCGAGGGCTCTCCACCTGTC 959
Qy 1050 CTCGCGCGGTGTGGGTACTTTCACAAGGGCGCGGTGAGAGCTCAGACCGCGCTCAGAC 1109
Db 960 GGGCTTCGACGCGGCTGCGCTATGGTATCGTTGGACGCGCCCGCAGAACCTTGAAGACT 1019
Qy 1110 GSGCGCCAGAGTGCACACCCACAGAGTCTGCGGATCTCTGTCGCGCTGATGCTGCT 1169
Db 1020 GTACAGCGCTCGCGCGGAGGCTTCGCGAGCGAGTGAAGAACCCGATCATGCTGCG 1079

QY 1170 GATCGCGGAGCCTTGCAGCAGATACCTCGGCATCCACTACGAGCGCAGAACCT 1229
Db 1080 CACCTACGCACTCTCG---GCGGCTACTACGATGCCATTACCTGCAGGCTCAGAAGAT 1136
QY 1230 GCGGCTGAGCTCGCAGCAGATGAGCGCGTCTCCAGAACCGGCTGACCTGCTGAC 1289
Db 1137 TCGCGGCTGATCAAGAACGACTTCTGTCAGCGCTTTGCCGAAGTGAAGCTGATCTCCGG 1196
QY 1290 CCGGACACGCTTACCTTTCACACGAGCTGTGAGCGGTCCGCAAGACACCATGTCCAT 1349
Db 1197 CCGGACACGCGGACCGGCTGGAAGATCGCGGAGAGAACGACGACCGGTTTCCCA 1256
QY 1350 GATCCACGATGACGGCAATGCGATCTCAACAGTGCCCGCTGAGACCTACCGGTCA 1409
Db 1257 GTACCTG-----GAAGACATCTACACCATCAACCGCAACCTCGCGGCT 1301
QY 1410 CCGCGGCTGACGGTCCACCGGCTGCGGGGAGAGGCGCTGCGGCTGCGCTCCCAAGT 1469
Db 1302 GCCGGGCTGTCCATGCGCGCGGCTTCTGTCG---ACGGCTGCGGCTGCGGTCCAGTT 1358
QY 1470 GATAGCGCGCACTTCGAGGA 1490
Db 1359 GCTCGCGGCTACTTCCAGGA 1379

RESULT 9

US-10-282-122A-30488
; Sequence 30488, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Lianguo
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malore, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Walli, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIORITY FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30488
; LENGTH: 1455
; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-10-282-122A-30488

Query Match 9.6%; Score 151; DB 13; Length 1455;
Best Local Similarity 49.0%; Pred. No. 2,1e-26;
Matches 598; Conservative 0; Mismatches 590; Indels 33; Gaps 6;
QY 273 CGCCACGAGGGGCACTGAGGACCTGACCCCGCGATCAAGGATCGCATCGCCATCGC 332
Db 189 CGCGAGAACGGCGCTCTCGCGCGCGCATCGCCACAAAGGACCTTTCTGACCCA 248
QY 333 CGGTATGCCACACACGAGGTCCTCGGATGTCCTCGCATGTCGCGCACGAGATGC 392
Db 249 GGGGTGACGACCGAGTGTGGTTCGAAGATGTCGCAACTTCGTCTGCCCTTACGACGC 308
QY 393 CGTGTGTGGAGCGGCTGTCTCGCGCAGGCGCCACCATCTGTCGGCAAGACCACTCGA 452
Db 309 CACCGTGTGCGAAGACTCACCGCCCGCGCGCTTACCTCTCGGCAAGCTGAACATGA 368
QY 453 GGACATGGCGATGGG---TATCGGTGAAGCGAGCGTCTACGGTCTCGCTGAAACCCGAA 509
Db 369 TGAATTCGCCCATGGGCTCGTGAACCACTCCAGCCACTACGGCGCGGTGAAGAACCCCTG 428
QY 510 CAACCCCGCCCAACGCGCAGCGGTGATCTTCCAGCGCTCCCGCGCTCCCTCGCTCCCG 569
Db 429 GAGCTTCGACCGGCTCGCGCGGCTCTCTCGCGGTTCGCGCGCGCAGTCTCGCGCGCG 488
QY 570 CATGTGCACTTGCCTCGGCTCGATGAGGACGAGCAGCATCCGATCCCGCGCGCATG 629
Db 489 CTTGCTGCCGCGCCACCGCGCACCGATACCGCGGCTCGATCCGCGCAACCGCGCGCGCT 548
QY 630 GTGCGGACTGTGCGCATGAAGCGCACCCACCGCTGTGCGGTCTTACGGCTTGACATA 689
Db 549 GACCAACCTCACCGGATCAAGCCAACTAGCGCGCGGTTCCCGCTGGGCGCATGATCGC 608
QY 690 CATGACACACCTTTGGAACACATCGGCGCCATACACAGGGGGTTCGAGTCAACGCCG 749
Db 609 CTACGCTTCCAGCTCGACACGCGCGCGCTGCGCGCACCGCGCGAGACTTCGCGCT 668
QY 750 GGTCTTCGAGGTGTGCGCGGCGGCGACTGCGCGGCGCTCAGTGGGTGCTAACTTCC 809
Db 669 GATGTGGGGTGATGCGCGGATTCGATCGAAGGACTCGACGCGT-----CGAACA 722
QY 810 GGAGCGGAGAACTACGGTTCGCGCTCGCGGAGGAGTATCGGTCTGAGATTCGCGGT 869
Db 723 GCGGTGAGCACTACTCTGCGCGCTCGAGAACCGCTGAGCGGCTTCGCGCATCGGCT 782
QY 870 CGTCGAGGAGTCACTGAGCGCGGAAAGCGTTCGCGCGCGCGGCGTATCGCGCTTCAACCA 929
Db 783 GCGCGGGAATACTTCGCGCGCGGCTTCGACACCGCGCTCGCGAGCGGTGCTGCGCGT 842
QY 930 GGCACTGGCGGCTCGAGAGCGCGGTCGACCATCGAGCGGCTCTCGGTGCGGTGTG 989
Db 843 GG---TCGAGGAGTGAAGACGCTCGGCGCGCACGGTGAAGGACATTTCCCTGCGGAACAT 899
QY 990 GACGCGGCTGCGCTATCCAGAGCGCGTATGCTTTCAACGCGCGGCTATGCGGGA 1049
Db 900 GCAGCAGCGCATCCCGGCTTACTACGTAATCGCGCGCGGCGGCGCTCTTCAACCTGTC 959
QY 1050 CTGCGCGGTGTGGGCTTACTTCACCAAGGGGCGCGTGGAGCTCAGCACCGCGCTTCAGCAC 1109
Db 960 CGGCTTCGACGGGTGGCTATGGCTATCGTTCGAGCGCGCGCGCGAGACCTTGAAGACCT 1019
QY 1110 GGGGCGCCAGAGTCGACACCAACCGATCTGGGATCTCTGTCGCGGTGATGCTGGT 1169
Db 1020 GTACAGCGCTCGCGCGGAGGCTTCGCGAGCGAAGTGAAGAACCGCATCTATGCTCGG 1079
QY 1170 GATCGCGGACCTCGCGGACGAAATACCTCGGCATCCACTACGCAAGCGCGAGACCT 1229
Db 1080 CACCTACGCACTCTCG---GCCGCTACTACGATGCGCTATCTACTGCGAGCTCAGAGAT 1136
QY 1230 GCGGCTGAGCTCGGCAAGCAGATCGACGCGCTCTTCCAGGACCGGCTCGATGTCAC 1289

Db 1137 TCGCCGGCTGATCAAGAACGACTTCGTACGGCCTTTGCCAAGTGGACGTGATCCTCGG 1196
Qy 1290 CCCGACACCGCTTACCGTTGCCAAGAGCTGTTGAGCGGTGCGCAAGACACCAATGTCAT 1349
Db 1197 CCCGACACCGCGAAGCCCGCTTGAAGATCGCGGAGAGAACGACCGGTTTCCCA 1256
Qy 1350 GATCCACGAGTACGCGGCAATGATCTCAACACAGTGCCTGACCTCACCGGTCA 1409
Db 1257 GTACCTG-----GGAAGACATCTACACCATCACGCCAACCTCGCCGGCCT 1301
Qy 1410 CCCGGCGCTGACCGTGCACAGGTGCGGCGAGAGAGGCTGCGCCGTTGGCCCTCAAGT 1469
Db 1302 GCGCGGGCTGTCCATGCGCCGCGCTTCGTG---ACGGCTGCGCGGTGTCCAGTT 1358
Qy 1470 GATAGGCGCGCACTTCGAGGA 1490
Db 1359 GCTCGCGCCTACTTCCAGGA 1379

RESULT 10

US-10-156-761-7248

; Sequence 7248, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; CURRENT FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 7248

; LENGTH: 1362

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1362)

US-10-156-761-7248

Query Match 9.0%; Score 142.6; DB 15; Length 1362;
Best Local Similarity 54.5%; Pred. No. 2e-24;
Matches 286; Conservative 0; Mismatches 239; Indels 0; Gaps 0;

Qy 286 CCACTGACGACCTGACCGCGCATCAGGACTGCATCGCCATCGCGGATGCCACC 345
Db 205 CCCCTGGCGGGGTGCCGTGCGCGTGAAGACAATCTCGCGGTACGGGGGAGTCCACT 264
Qy 346 ACGAACGGGTCCCGGATCTCCGCACTGTGATCGCCACCGAGGATGCGGTGGAG 405
Db 265 CGCCACGACTCGCGGCCACCCCGACACCCCGCGCCGACACCATGTACCGTGGCC 324
Qy 406 CGGCTGCTCGGCGCAGGCGCCACCATGTGCGGCAAGAGAACCTCGAGGACATGGCGATG 465
Db 325 CGGCTGCTCGGCGGCGGTGCGCTGCTGCGGGGTGACGAACGTGCCCGGAGCTTCGGTG 384
Qy 466 GGTATCGGTGAAGCAGCGGTCTACGGTCTCGGTGTAACCCGGAACAACCCCGCCACGGC 525
Db 385 TTGCGCACCGACGCGGCGTGCACGCGACGCGCCGCGACACCTCGCGGACG 444
Qy 526 ACGGTGATCTTCAGCGGCTCGCGCGCTGCGGTGCGCGGATGGTGCATCTTCGCC 585
Db 445 GCGGCGGCTCTCTCGGCGGGGAGCGCGCGCGGTGCGCGCGCGGATGGTCCCGATCGCG 504

Qy 586 CTGGCGCTCGATGAGGACGAGCATCCGGATCCCGGCCGCGCATGTCGGGACTGTCGCGC 645
Db 505 CTCGCAACAGAGGATGGGTCCCTCGCGATCCCGCGCCCAACTCGCGTCTGTCGCGC 564
Qy 646 ATGAGGCGACCCACGGCTGTCCTTTACCGCTTGAACATACATGAGACACACCTTG 705
Db 565 CTGAAGCCGGGTACGGACGGTCCGCGCGGCGCATCGAAACGGCGACTGCTCGGCATG 624
Qy 706 GACCATCGGGCCCATCACCGGGGGTTCAGCTCAACGCCCGGCTCTCGAGGTGTTG 765
Db 625 TCCGAGAACGGCCCGCTCGGACACACGGTTCAGGACGCGCCGCTGATTTCCGCTGCTG 684
Qy 766 GCGCGGGCGCATGTCGCGGACCCCTCAGTGGGTGGTGAACCTTCCG 810
Db 685 GCGCGGACGGTCAACGGCGTCCGCGGACACGAGGCGGATACGCCCG 729

RESULT 11

US-10-282-122A-11356

; Sequence 11356, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 11356

; LENGTH: 1488

; TYPE: DNA

; ORGANISM: Burkholderia cepacia

US-10-282-122A-11356

Query Match 9.0%; Score 141.4; DB 13; Length 1488;
Best Local Similarity 48.8%; Pred. No. 3.9e-24;
Matches 527; Conservative 0; Mismatches 521; Indels 33; Gaps 4;
Qy 266 TGAAGCGCGCCACGGAGGGGCCACTGAGCGACTGACCGCCGCGATCAAGGACTGCATCG 325

Db 176 TCGCGCGCGCGCGCGCGCGCTGACCGGCTGCCGATCGGCACAGAGCGTGTTCG 235
QY 326 CCAATCGCCGGTATGCCACCAAGACGGGTCCGGATGCTCCGACTGTGATCGCACCG 385
Db 236 TACAGCGCGCGTGGCGCTCGACCGCGCGGTTCGAAGATGCTCGCAATACGAGAGCCGCT 295
QY 386 AGGATCCGTGTGTGGAGCGGCTCTCGCGCGAGCGCCACCATCGTCGGCAAGACA 445
Db 296 TCGACGGACCGTCTGTGCGCCGCTGACGCGCGCGCATGTGCTCGCTCGCAAGACA 355
QY 446 ACTCGAGGACATGGCGATGGGATTCGTTGAAGGA---GCTTACGGTCTCTGCGCTGA 502
Db 356 ACATGACGAGTTCGCGATGGGCTGCTCGAAGAGAAATCCGCGTTCGGCGCGTGAAGA 415
QY 503 ACCCGAAACCCCGCCACGCGACGCGGTGGATCTTCCAGCGGCTCCGCGCTGCGCTCG 562
Db 416 ACCGTGGACAGACGAGCGGTGCGCGCGCGAGCTCGCGCGCGAGCTCGCGCGCGCTCG 475
QY 563 CTGCGCGCATGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 622
Db 476 CCGCGCGCTCG 535
QY 623 CCGCATGTGCGGACTGCTCGCATGAAGCGGACCGCGCGCTGCTGCGCTTACGGCC 682
Db 536 GTCTGTTCCGCGCGGTGACGCGGATCAAGCGGACCTACGCGCGCGGTGCTGCGCTACGGGA 595
QY 683 TGACATATGACACACACTTGGACCATCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 742
Db 596 TGATCGCTGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 555
QY 743 AGCGCGGCTCTCGAGGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 797
Db 656 GCGCGCTGCTGCTGAAACGCGATGGCGGTTGCGAGCGGTGACTGCGAGCGCTCGAGC 715
QY 798 -----GCGTAACTTCCGAGCGCGGAGAACTACGCGCTCGCGCTCG 838
Db 716 GCGACGACGAGACTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 775
QY 839 GCGAGGAGTATCCGCTGAGATTCGCGGTGCTGAGGAGTCACTGGAGCGCAACGCTG 898
Db 776 GCAAGCGCTGCGCGCGCTGCGCATCGGCTTCCGAAACAGTATTTCCGCGC---CGGCC 832
QY 899 CGACGCGGACGCTGATCGCGCGCTTCAACGAGGACTGCGCGCTCGAGAGCGCGGTG 958
Db 833 TCGCCACGAGTTCG 892
QY 959 CGACCATCGAGCGGCTCTCGGTGCGCTTGTGAGCGCGCGCTGCGCTATCCAGAGCGCG 1018
Db 893 CGACGCTGCTGCGCGGTGCTGCTGCGCGAGAGCTGCTGATCCCGGTGACTAGTGA 952
QY 1019 TGATGCTTTCAACGCGCGCGCTATGCGGACTTCGCGGTGCTGCTTCCAGAGG 1078
Db 953 TCGCGCGCGGAGCGTCTGCGAACTGTCGCGCTTCAGCGCGGTGCGTTTCGCGCAC 1012
QY 1079 GCGCGGTGACGCTGACGCGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1138
Db 1013 GCGCGCGCGAGTACGCGCGCTGCTGACATGATACAGAGTCTGCGCGCGCGAGGCTCG 1072
QY 1139 ATCTGCGATCTCTGCTCGCGCTGATCTGCTGATTCGCGAGCACCTGCGCGCGAGTACC 1198
Db 1073 GGCTGAAGTGAAGCGCG---GATCTGCTGCGCGCGCTGATGCTGCTGCTGCTGCT 1129
QY 1199 TCGGCTTCCACTACGCGAGCGCGAGAACTGCGGCTGGAGCTTCGCGAGAGATTCGAG 1258
Db 1130 ACAGCGGTACTTACTTCAGCGCGCGAGAGATCCGCGCGCATCTCGGAGCATTTCCAG 1189
QY 1259 CCGTCTTCAGGACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1318
Db 1190 AAGCGTTCAAGTCTCGAGGTGATCATGGGCGCGCGCGCTGCGCGCGCGCTGCGATG 1249
QY 1319 T 1319
Db 1250 T 1250

RESULT 12

US-10-282-122A-15096
; Sequence 15096, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 15096
; LENGTH: 1536
; TYPE: DNA
; ORGANISM: Bordetella pertussis
US-10-282-122A-15096

Query Match 9.0%; Score 141.4; DB 13; Length 1536;
Best Local Similarity 53.9%; Pred. No. 3.8e-24;
Matches 314; Conservative 0; Mismatches 266; Indels 3; Gaps 1;

QY 266 TGGAGGGCGCCACGAGGGGCGCACTGAGCGACCTGAGCGCGCGGATCAAGGACTGCATCG 325
Db 197 TGGCGCGCGGACGCGCGCGCGCGCGCTGCGCGGCGATCCCGCATCGCCCAAGAGCGCTTCG 256
QY 326 CCATCGCGCGGTATGCCCGACCAAGCGGTCGCCGATGCTCCCGACTGCTGATCGCCACCG 385
Db 257 TCACCGGTGGCTGGCGCACCGACCGCAGCAGCAAGATGCTGGCGGCTATGCCAGCGCT 316
QY 386 AGGATGCGGTGGTGGAGGGGCTGCTCGGGCAGCGGCGCCACCATCGTTCGGCAGAGCA 445
Db 317 TCGACGCGCACCGTGGTTCGAGCGCTTGGCGGCGCGCGGTATCGCTGGGCAAGTGA 376
QY 446 ACCTCGAGGACATGGCGATGGGTATCGGTGAAGGACG---GTCTAGGCTCTCGCTGA 502
Db 377 ACTCGGATGAGTTCGCCATGGCTCGGCGCAACGAGAACTCGGCGCTACGGCCCGGTGCGCA 436
QY 503 ACCGAAACCCCGCGCCACGCGACCGGCTGATCTTCAGGGGCTCCGGCGCTCGCGCTCG 562

